8558392

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

January 06, 2025

THIS IS TO CERTIFY THAT ANNEXED IS A TRUE COPY FROM THE RECORDS OF THIS OFFICE OF THE FILE WRAPPER AND CONTENTS OF:

APPLICATION NUMBER: *14/512,414*
FILING DATE: *October 11, 2014*
PATENT NUMBER: *9350649*
ISSUE DATE: *May 24, 2016*

Certified by

Acting Under Secretary of Commerce
for Intellectual Property
and Acting Director of the United States
Patent and Trademark Office

PTO/SB/81 (01-09)
Approved for use through 11/30/2011. OMB 0651-0035
U.S. Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

POWER OF ATTORNEY OR REVOCATION OF POWER OF ATTORNEY WITH A NEW POWER OF ATTORNEY AND CHANGE OF CORRESPONDENCE ADDRESS

Application Number	
Filing Date	2014-10-11
First Named Inventor	ERIC DELANGIS
Title	MULTIPATH COMMUNICATION DEVICES AND METHODS
Art Unit	
Examiner Name	
Attorney Docket Number	ED 2000

I hereby revoke all previous powers of attorney given in the above-identified application.

☐ A Power of Attorney is submitted herewith.

OR

☒ I hereby appoint Practitioner(s) associated with the following Customer Number as my/our attorney(s) or agent(s) to prosecute the application identified above, and to transact all business in the United States Patent and Trademark Office connected therewith:

64828

OR

☐ I hereby appoint Practitioner(s) named below as my/our attorney(s) or agent(s) to prosecute the application identified above, and to transact all business in the United States Patent and Trademark Office connected therewith:

Practitioner(s) Name	Registration Number

Please recognize or change the correspondence address for the above-identified application to:

☒ The address associated with the above-mentioned Customer Number.

OR

☐ The address associated with Customer Number:

OR

☐	Firm or Individual Name				
Address					
City		State		Zip	
Country					
Telephone		Email			

I am the:

☒ Applicant/Inventor.

OR

☐ Assignee of record of the entire interest. See 37 CFR 3.71
Statement under 37 CFR 3.73(b) (Form PTO/SB/96) submitted herewith or filed on _____

SIGNATURE of Applicant or Assignee of Record

Signature		Date	10/11/14
Name	ERIC DELANGIS	Telephone	970-310-3576
Title and Company	INVENTOR		

NOTE: Signatures of all the inventors or assignees of record of the entire interest or their representative(s) are required. Submit multiple forms if more than one signature is required, see below*.

☐ *Total of _____ forms are submitted

This collection of information is required by 37 CFR 1.31, 1.32 and 1.33. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 3 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

RED HAT 1010
RED HAT V. CAS

Electronic Patent Application Fee Transmittal	
Application Number:	
Filing Date:	
Title of Invention:	MULTIPATH COMMUNICATION DEVICES AND METHODS
First Named Inventor/Applicant Name:	Eric Delangis
Filer:	Harold C. McGurk
Attorney Docket Number:	ED 2000
Filed as Small Entity	

Utility under 35 USC 111(a) Filing Fees

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Utility filing Fee (Electronic filing)	4011	1	70	70
Utility Search Fee	2111	1	300	300
Utility Examination Fee	2311	1	360	360
Pages:				
Claims:				
Independent Claims in Excess of 3	2201	1	210	210
Miscellaneous-Filing:				
Petition:				

RED HAT 1010
RED HAT V. CAS

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				
Miscellaneous:				
			Total in USD ($)	940

RED HAT 1010
RED HAT V. CAS

Electronic Acknowledgement Receipt	
EFS ID:	20393175
Application Number:	14512414
International Application Number:	
Confirmation Number:	9530
Title of Invention:	MULTIPATH COMMUNICATION DEVICES AND METHODS
First Named Inventor/Applicant Name:	Eric Delangis
Customer Number:	64828
Filer:	Harold C. McGurk
Filer Authorized By:	
Attorney Docket Number:	ED 2000
Receipt Date:	11-OCT-2014
Filing Date:	
Time Stamp:	16:19:40
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	Credit Card
Payment was successfully received in RAM	$940
RAM confirmation Number	6093
Deposit Account	
Authorized User	

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)

RED HAT 1010
RED HAT V. CAS

1	Application Data Sheet	ED_2000_ADS_SB14.pdf	1023461 8c1e503610b23535b442027e000b9ce52072791f	no	4
Warnings:					
Information:					
2	Specification	ED_2000_Specification.pdf	128315 93d1630b034c61e99f4083ab977506a8c9365290	no	24
Warnings:					
Information:					
3	Claims	ED_2000_Claims.pdf	29140 ac08d8eb4abe9059ef532f79a61687fb3df42042	no	4
Warnings:					
Information:					
4	Abstract	ED_2000_Abstract.pdf	23358 3222d6135b51c50b10d23188bdbf20b84a1b09b6	no	1
Warnings:					
Information:					
5	Drawings-only black and white line drawings	ED_2000_Drawings_1-9.PDF	2620724 5f2cabc1b03f19772e97df396e0ff589f976b	no	9
Warnings:					
Information:					
6	Oath or Declaration filed	ED_2000_Declaration.pdf	3007229 c1e7741d95a0bb70590719cc99d817d8b03d2235	no	1
Warnings:					
Information:					
7	Power of Attorney	ED_2000_Power_Attorney.pdf	3142157 0862134f9404a4582cbec0f3194c6779f50d94f9	no	1
Warnings:					
Information:					
8	Fee Worksheet (SB06)	fee-info.pdf	36291 7c2a39e4f3c86f5148a965af73b5fb918134487f3c	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	10010675		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

RED HAT 1010
RED HAT V. CAS

PTO/SB/14 (11-08)
Approved for use through 01/31/2014. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	ED 2000
	Application Number	
Title of Invention	MULTIPATH COMMUNICATION DEVICES AND METHODS	

The application data sheet is part of the provisional or nonprovisional application for which it is being submitted. The following form contains the bibliographic data arranged in a format specified by the United States Patent and Trademark Office as outlined in 37 CFR 1.76.
This document may be completed electronically and submitted to the Office in electronic format using the Electronic Filing System (EFS) or the document may be printed and included in a paper filed application.

Secrecy Order 37 CFR 5.2

☐ Portions or all of the application associated with this Application Data Sheet may fall under a Secrecy Order pursuant to 37 CFR 5.2 (Paper filers only. Applications that fall under Secrecy Order may not be filed electronically.)

Applicant Information:

Applicant 1 [Remove]

Applicant Authority ◉ Inventor ○ Legal Representative under 35 U.S.C. 117 ○ Party of Interest under 35 U.S.C. 118

Prefix	Given Name	Middle Name	Family Name	Suffix
	Eric	M.	Delangis	

Residence Information (Select One) ◉ US Residency ○ Non US Residency ○ Active US Military Service

City	Huntsville	State/Province	AL	Country of Residence i	US

Citizenship under 37 CFR 1.41(b) i: US

Mailing Address of Applicant:

Address 1	231 Brevard Blvd		
Address 2			
City	Huntsville	State/Province	AL
Postal Code	35811	Country i	US

All Inventors Must Be Listed - Additional Inventor Information blocks may be generated within this form by selecting the **Add** button. [Add]

Correspondence Information:

Enter either Customer Number or complete the Correspondence Information section below.
For further information see 37 CFR 1.33(a).

☐ **An Address is being provided for the correspondence Information of this application.**

Customer Number	64828		
Email Address	claymcgurk@gmail.com	[Add Email]	[Remove Email]

Application Information:

Title of the Invention	MULTIPATH COMMUNICATION DEVICES AND METHODS		
Attorney Docket Number	ED 2000	Small Entity Status Claimed	☒
Application Type	Nonprovisional		
Subject Matter	Utility		
Suggested Class (if any)		Sub Class (if any)	
Suggested Technology Center (if any)			
Total Number of Drawing Sheets (if any)	9	Suggested Figure for Publication (if any)	7

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	ED 2000
	Application Number	
Title of Invention	MULTIPATH COMMUNICATION DEVICES AND METHODS	

Publication Information:

☐ Request Early Publication (Fee required at time of Request 37 CFR 1.219)

☐ **Request Not to Publish.** I hereby request that the attached application not be published under 35 U.S. C. 122(b) and certify that the invention disclosed in the attached application **has not and will not** be the subject of an application filed in another country, or under a multilateral international agreement, that requires publication at eighteen months after filing.

Representative Information:

Representative information should be provided for all practitioners having a power of attorney in the application. Providing this information in the Application Data Sheet does not constitute a power of attorney in the application (see 37 CFR 1.32). Enter either Customer Number or complete the Representative Name section below. If both sections are completed the Customer Number will be used for the Representative Information during processing.

Please Select One:	◉ Customer Number	○ US Patent Practitioner	○ Limited Recognition (37 CFR 11.9)
Customer Number	64828		

Domestic Benefit/National Stage Information:

This section allows for the applicant to either claim benefit under 35 U.S.C. 119(e), 120, 121, or 365(c) or indicate National Stage entry from a PCT application. Providing this information in the application data sheet constitutes the specific reference required by 35 U.S.C. 119(e) or 120, and 37 CFR 1.78(a)(2) or CFR 1.78(a)(4), and need not otherwise be made part of the specification.

Prior Application Status	Patented			Remove	
Application Number	Continuity Type	Prior Application Number	Filing Date (YYYY-MM-DD)	Patent Number	Issue Date (YYYY-MM-DD)
	Continuation of	13531294	2012-06-12	8861349	2014-10-14
Prior Application Status	Patented			Remove	
Application Number	Continuity Type	Prior Application Number	Filing Date (YYYY-MM-DD)	Patent Number	Issue Date (YYYY-MM-DD)
13531294	Continuation of	12581852	2009-10-19	8228801	2012-07-24
Prior Application Status	Patented			Remove	
Application Number	Continuity Type	Prior Application Number	Filing Date (YYYY-MM-DD)	Patent Number	Issue Date (YYYY-MM-DD)
12581852	Continuation of	10686375	2003-10-14	7606156	2009-10-20
Prior Application Status	Expired			Remove	
Application Number	Continuity Type	Prior Application Number	Filing Date (YYYY-MM-DD)		
10686375	non provisional of	60418521	2002-10-15		

Additional Domestic Benefit/National Stage Data may be generated within this form by selecting the **Add** button. Add

Foreign Priority Information:

PTO/SB/14 (11-08)
Approved for use through 01/31/2014. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	ED 2000
	Application Number	
Title of Invention	MULTIPATH COMMUNICATION DEVICES AND METHODS	

This section allows for the applicant to claim benefit of foreign priority and to identify any prior foreign application for which priority is not claimed. Providing this information in the application data sheet constitutes the claim for priority as required by 35 U.S.C. 119(b) and 37 CFR 1.55(a).

Remove

Application Number	Country [i]	Parent Filing Date (YYYY-MM-DD)	Priority Claimed
			(•) Yes ( ) No

Additional Foreign Priority Data may be generated within this form by selecting the **Add** button. Add

Assignee Information:

Providing this information in the application data sheet does not substitute for compliance with any requirement of part 3 of Title 37 of the CFR to have an assignment recorded in the Office.

Assignee 1 Remove

If the Assignee is an Organization check here. ☐

Prefix	**Given Name**	Middle Name	**Family Name**	Suffix

Mailing Address Information:

Address 1			
Address 2			
City		**State/Province**	
Country [i]		Postal Code	
Phone Number		Fax Number	
Email Address			

Additional Assignee Data may be generated within this form by selecting the **Add** button. Add

Signature:

A signature of the applicant or representative is required in accordance with 37 CFR 1.33 and 10.18. Please see 37 CFR 1.4(d) for the form of the signature.

Signature	/H C McGurk, Reg. 34964/			Date (YYYY-MM-DD)	2014-10-11
First Name	Harold	Last Name	McGurk	Registration Number	34964

This collection of information is required by 37 CFR 1.76. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 23 minutes to complete, including gathering, preparing, and submitting the completed application data sheet form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these records.

2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.

3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.

4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).

5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.

6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).

7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.

8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.

9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

MULTIPATH COMMUNICATION DEVICES AND METHODS

RELATED APPLICATIONS

[0001] This application is a continuation of U.S. Patent Application Serial No. 13/531,294 filed June 12, 2012, which is a continuation of U.S. Patent Application Serial No. 12/581,852 filed October 19, 2009, which is a continuation of U.S. Patent Application Serial No. 10/686,375 filed October 14, 2003, which claims priority to U.S. Provisional Application Serial No. 60/418,521, filed October 15, 2002, all of which are incorporated by reference herein.

BACKGROUND OF INVENTION

[0002] The primary function of the RCG is to provide a platform for Competitive Local Exchange Companies (CLECs) to compete with the Local Exchange Companies (LECs) for residential voice and data services. The LECs have a strangle hold on the "last mile" of copper over which almost all residential customers receive their communications services. Today, most CLECs resell the existing services that are offered by the LECs. CLECs make their money by purchasing these services from the LECs at a discount and marking them up to the customer. Since these services are regulated and controlled by the Public Utilities Commission (PUC) the pricing for these services to the CLECs is set. The monthly CLEC discount is approximately $5 to $6 per line per month for basic rate services. This means, in practicality, that a CLEC can only charge an additional $5 or $6 to the residential customer for basic rate services otherwise they would be more expensive than the LEC, and no one would buy the services from them; telephone customers would simply stay with the LEC.

[0003] To be competitive then, CLECs try to sell additional services such as call waiting, caller ID, etc. where they have as much as a 50% margin. Some CLECs also bundle long distance services with their packages as additional incentive to get residential customers to buy local telephone service from them. Since long distance fees are as low as pennies per minute to the customer, there is little profit in long distance for the CLECs. As it is, the pricing structure is such that the LECs will stay in control of the residential market, and the CLECs are nothing more than a marketing arm of the LEC. This structure also keeps the CLECs at an equal competitive advantage, or disadvantage, with respect to each other. The only thing that one

RED HAT 1010
RED HAT V. CAS

CLEC could do to gain more customers is to be a better marketier. Since the CLECs are unable to discount their services because their margins are too slim as it is, the net effect is that the consumer looses by not really getting competition in the residential markets, and prices are kept artificially high.

[0004] For the CLECs to be truly competitive in the local telecommunications markets, they must be able to do more than simply resell the LECs services. CLECs must be able to offer many more features or services than those that the LEC provides, and they must be able to do this at a lower cost, in order to attract new customers.

[0005] So, what new services can the CLECs offer? As things stand today there are none, with the exception of DSL or Cable services which require tremendous infrastructure costs on behalf of the CLEC. However, by deploying an RCG in a residence, CLECs can offer advanced new calling features and services over the existing POTS line. Their costs for the POTS line from the LEC remains the same, but they can now charge for high bandwidth data services as well as additional calling and messaging features that the LECs cannot offer. This is their competitive advantage.

Competition:

[0006] Many companies are currently trying to provide VoIP services to residential customers over high-speed or broadband backbones such as DSL and cable modems, over Internet connections via a computer. The high-speed services require expensive infrastructure enhancements to the local telephone companies' facilities and cable plants, as well as expensive modems at the customer's end. These services are also difficult to deploy and too complicated for the average user. Internet based VoIP services suffer from Internet delays caused by a lack of true Quality of Service (QOS) capabilities across the Internet. Existing VoIP services also require access to a computer and a technical knowledge to be able to make it work. Because of these problems, VoIP services are primarily targeted at the corporate market place where private IP networks that can carry VoIP services with a true guaranteed QOS exist, or can be controlled.

[0007] The only available alternatives for the residential user for high speed services are DSL and cable modems. DSL and Cable based communications reach only a small segment of the population, and growth has failed to live up to initial expectations. The recent collapse of the

internet boom has led to the failure of many DSL service providers and has effectively eliminated DSL as a potential service for most POTS users for the immediate future.

SUMMARY OF THE INVENTION

[0008] The RCG combines the capabilities and features of IP routers, Class 5 circuit switches and wireless LANs in a small stand-alone external modem like device. Additional capabilities for a hands-free speakerphone device with a large touch display are also detailed. The RCG can employ a large display with sufficient resolution to provide a functional user interface to access all of the advanced features offered by the RCG, but also to be able to provide video telephone services.

[0009] The device is placed in a residence and connected to the primary POTS line coming from the LEC. Up to three standard telephones or FAX machines (or any combination) are typically connected into the RCG. It also has a Universal Serial Bus (USB) and/or a 10-100 Ethernet interface for connection to a computer or a Local Area Network (LAN). In addition, it has an 802.11b/g wireless interface for wireless home networking as well as broadband-to-the-home services.

[0010] The system is installed in a customer's home by simply plugging it into the existing incoming telephone line that provides POTS services from the LEC. One or more standard telephones are then plugged into its three line-output jacks. This is the simplest and most basic usage. In this implementation, the user only is interested in gaining additional telephone lines into the household. No computer connection is required in order to use the system. However, if data communications are desired, then there are three additional interfaces for computer connectivity. Typically either an Ethernet port, USB port or firewire (IEEE 1394) port is used to connect to a PC or LAN. The 802.11b/g interface is for wireless communications between the RCG and PCs, other RCGs as well as any other types of compatible wireless devices. A version of the RCG includes a speakerphone and video camera to provided for video telephone services.

Services provided by the RCG:

[0011] The following discussion refers to figure 1. The RCG has a main system processor that is responsible for all house keeping and traffic routing tasks. The Main CPU 19 routes data

to and from the Ethernet Port 11, USB Port 12, Modem/DAA 41, DSP Engine 33, 802.11b/g Module 36 and the Display and Keypad 10.

[0012] The RCG provides a total telecommunications communications services package for residential and small business customers. With the RCG, a CLEC can offer its customers dialtone for up to three separate telephone numbers per residence (the limit of three telephone numbers is an arbitrary number based upon typical residential usage and is not a limit imposed by the system), all local and long distance services, 24/7 always-on Internet access, a host of custom messaging and calling features, as well as Broadband access and lifeline service. All of these services can be offered over a customer's existing POTS line, and it doesn't require any network infrastructure build out like DSL and Cable services. Because there is no expensive network equipment required, it can be readily deployed throughout the country to expand the CLECs customer base quickly.

[0013] The RCG is designed to enhance, not replace, the existing telephone line, enabling it to provide new combined packetized voice and data services. Meanwhile, all existing telecommunications infrastructure and equipment will continue to work seamlessly with the RCG. Aside from simply installing the low cost RCG device in the home telephone wiring, no other equipment needs to be purchased by either the home owner or telephone companies. This eliminates the tremendously expensive infrastructure costs for communications companies that are trying to provide voice services using DSL or cable modems.

The RCG VIP Home Gateway features:

[0014] Up to three telephone lines i.e. numbers, over a single POTS line and/or wireless connection.

[0015] Ethernet, USB and IEEE 1394 computer interfaces.

[0016] Full G1, G2 and G3 FAX support.

[0017] Full support for all CLASS services.

[0018] Broadband networking for home computers and appliances as well as wide area networking using a combination of a plurality POTS lines connected to a plurality of RCGs and/or a direct wireless connection to a wireless access point. (These interfaces allow smart devices to communicate to other devices within the home, or over the Internet through the RCG.)

[0019] Automated, remote upgradability and configuration. (Requires the customer to do nothing.)

[0020] Transparent user operation. (User doesn't have to know that any new technology is being employed)

[0021] 100% compatible with existing public telephone system.

[0022] Video Telephone capabilities.

[0023] No local or long distance usage charges.

[0024] No Internet access fees.

[0025] Lower monthly communication costs.

[0026] Multiple telephone calls at the same time.

[0027] Voice messaging, follow me features, call forwarding, automatic call back, conferencing, custom call routing. Home networking capabilities, all CLASS calling features, high speed broadband networking, continuous 24/7 internet access. and many others.

[0028] All of these features are available even while connected to the Internet.

[0029] Very low cost of entry into the market, especially as compared to broadband alternatives.

[0030] New calling features that are not currently available to home users.

BRIEF DESCRIPTION OF THE DRAWINGS

[0031] Figure 1 shows the RCG functional block diagram with the preferred embodiment of an 802.11 wireless connection.

[0032] Figure 2 shows the network topology for native RCG to RCG communications.

[0033] Figure 3 shows the network topology for an RCG to PSTN call.

[0034] Figure 4 shows the RCG top view with associated functions.

[0035] Figure 5 shows the RCG rear view and associated connections.

[0036] Figure 6 shows a typical application with peripheral equipment connections.

[0037] Figure 7 shows an ad-hoc wireless LAN configuration with ppp multilink data distribution and wireless connectivity.

[0038] Figure 8 shows an ad-hoc wireless LAN configuration with ppp multilink data distribution and wireless connectivity with connection to a broadband neighborhood access point (Wi-Fi hot spot).

[0039] Figure 9 shows an ad-hoc wireless LAN configuration with ppp multilink data distribution and wireless connectivity with wireless data hopping with POTS connections through numerous different Class 5 offices.

DETAILED DESCRIPTION OF THE INVENTION

[0040] The preferred embodiment is detailed herein using the ubiquitous 802.11 wireless protocol. It is by no means insinuated that this is the only wireless scheme that can be used by this device; rather, it is intended that any wireless transmission scheme can be used for wireless communications between this device and other wireless devices as various applications may dictate.

System Initialization

[0041] Refer to figures 1, 2 and 3 for the following discussion. When the RCG is first plugged into an AC power outlet the Main CPU 19, performs a Power-on Self Test (POST) routine. Once the POST completes and the RCG is connected to an active POTS line, connected to the LEC, by means of the Incoming POTS Port 40, the RCG automatically initiates a modem connection via the Modem/DAA 41, to a registration service via a toll free 800 telephone number. The registration service connects directly to the service provider's sales or customer service offices via a VoIP connection. An operator or an automated service answers the call and sets up a new account for the customer. Once the account is set up, the service provider's network automatically sends control information to the device to store the required routing and feature/service information in its non-volatile memory of its Main CPU 19. As part of this information, a local dialup telephone number is sent to the device that it will use for local network access in the same fashion as a standard ISP connection is made from a home computer. The URL of the Softswitch/SIP Proxy Server and Registrar 51 that will be controlling this particular device is also stored at this time. During the service activation and configuration process, the customer service operator can send test requests to the device, as well as call into it on another telephone number (assuming an additional telephone number is part of the users class of service) or the speakerphone to verify that it is configured and working properly. The final step of the configuration and initialization process is to register the device with the appropriate

RED HAT 1010
RED HAT V. CAS

Softswitch/SIP Proxy Server and Registrar 51 so that the Softswitch/SIP Proxy Server and Registrar 51 knows that it is the controlling SIP Proxy for this RCG. Also configured at this time is the username/password verification server (this is a standard network device and is not shown) that is used to allow the RCG's modem to make a connection onto the service provider's network. Once the RCG configuration is completed, it disconnects the call by hanging up the Modem/DAA 41, then dials the new local network access number that had been sent during the configuration process and establishes a constant local network connection. This connection is maintained continuously on a 24/7 basis, and is immediately reestablished whenever it drops. Should the RCG become unable to reestablish this connection after several tries, it will switch over to failsafe mode and inform the user via the Display and Keypad 10. The RCG will try to reestablish the connection every few minutes until it finally succeeds.

[0042] By default, the first telephone port, POTS 1 Port 30, is assigned as the termination port for the existing POTS service that is connected to Incoming POTS Port 40. This is the existing telephone service that the user already has as his primary telephone service from his LEC. POTS 1 Port 30, typically has connected to it a standard telephone set that does not require any power (other than that derived over the POTS interface) to operate properly. POTS 1 Port 30 is the Lifeline telephone port for the user. In the case that all power is lost to the RCG, POTS 1 Port 30 will fail to a direct connection to the Incoming POTS Port 40 telephone line that comes into the RCG from the LEC to provide uninterrupted telephone services to the user. In this state, however, none of the enhanced VoIP services of the RCG are available to the user since the user simply has a stand POTS telephone service to the LEC.

[0043] Two additional and unique telephone numbers can be assigned to the other two telephone ports, POTS 2 Port 29 and POTS 3 Port 28, by the service provider at any time. These numbers are drawn from a pool of available telephone numbers assigned to the service provider. POTS 2 Port 29 and POTS 3 Port 28 must to have standard telephone numbers assigned to them so that the PSTN is able to route calls from its network to these ports. For calls originating from IP based telephones, these numbers are not necessarily needed since POTS 1 Port 30, POTS 2 Port 29 and POTS 3 Port 28 are uniquely identifiable to IP networks via their IP addresses assigned through the Dynamic Host Configuration Protocol (DHCP) at the time that the RCG establishes a connection to the service provider's network. Since POTS 2 Port 29 and POTS 3 Port 28 must have a telephone number assigned to them, it is a matter of convenience to use the

assigned telephone numbers as part of their SIP address for IP routing of their data. In this way, all three user POTS ports, POTS 1 Port 30, POTS 2 Port 29 and POTS 3 Port 28 are all addressed the same manor.

[0044] At this point, the user is able to take advantage of all of the RCG's features that have been activated according to his service agreement. The user can at any time connect additional telephones, a computer or a computer network to the RCG. The user will only be able to use those telephone ports (POTS 1 Port 30, POTS 2 Port 29 or POTS 3 Port 28) that have been activated, and assigned unique telephone numbers by the service provider. Any type of standard telephone can be connected to the three POTS ports, POTS 1 Port 30, POTS 2 Port 29 and POTS 3 Port 28. In the event that a cordless telephone or a telephone that derives its power from an AC outlet is connected to POTS 1 Port 30, the lifeline functionality of this port will be ineffective since the telephone itself will not operate once power is lost.

[0045] The speaker phone feature of the RCG, comprised of Speakerphone Drivers 22, Microphone 24, Speaker 25 and DSP Engine 33, can be used on any of the activated telephone numbers and is by default enabled to operate on POTS 1 Port 30. Once either of the other two telephone ports, POTS 2 Port 29 or POTS 3 Port 28, is activated, the speakerphone function is activated for those lines as well. To use the speakerphone for any assigned telephone number, the user simply presses either the Line Select Button 79-81 for Line 1, Line2 or Line 3 on the face of the RCG, and the speakerphone function is activated on that line. This is convenient for answering any call directly at the RCG. The RCG may or may not have a number keypad for originating calls. This feature may be added or removed based upon the target market price points.

[0046] The determination as to which Inter Exchange Carrier (IXC) the RCG will route its data through is determined by any of a number of criteria including, but not limited to, whether the RCG is a private labeled product, or which carrier has the lowest cost routing, or if a specific carrier is specified by the customer, or and number of other considerations. The RCG can be programmed to change its routing path to any IXC at any time by the service provider. Complicated routing algorithms can be programmed into the RCG to control which IXC it will use at any time.

Failsafe Mode

[0047] During any time that the RCG is in failsafe mode i.e. Lifeline failsafe logic 39 is activated, the RCG will only be able to receive or initiate telephone calls via the telephone connected to its primary telephone port, POTS 1 Port 30 or through its speakerphone feature. In this condition, all activated telephone numbers are ported via the PSTN to the primary (first) telephone number for delivery to the user's telephone that is connected to POTS 1 Port 30. While in failsafe mode, the modem is not activated, and the RCG appears to the local Class 5 office as a standard telephone, and the PSTN will route all incoming calls to the RCG as it normally would to any standard telephone.

Call Flow 1:

[0048] Dialed call origination from a telephone connected to an RCG POTS port and terminating at another user's RCG POTS port:

[0049] When a user picks up a telephone connected to any of the three POTS ports, POTS 1 Port 30, POTS 2 Port 29 and POTS 3 Port 28, an off hook condition is detected by the SLIC/CODEC Interface 27 that is connected to the DSP Engine 33 by way of the DSP Audio Interface 31. The DSP Engine 33 responds by sending digitally generated dialtone to the appropriate POTS interface over the DSP Audio Interface 31. (Note: all BORSCHT telephony functions are performed by the DSP Engine 33 and the SLIC/CODEC Interface 27.) Upon hearing the dialtone, the user enters the desired telephone number on his telephone keypad. The Dual Tone Multi-Frequency (DTMF) digits that the user dials are decoded within the DSP Engine 33. The dialed digits are transferred to the Main CPU 19 via the CPU/DSP Communications Path 34. The Main CPU 19 polls the DSP Engine 33 approximately every two milliseconds looking for data. The Main CPU 19 polling rate varies according to the process that is being performed. The CPU/DSP Communications path 34 also can generate an interrupt to force data transfer between it and the Main CPU 19.

[0050] The Main CPU 19 generates a SIP Invite IP packet using the dialed digits as the called destination address and the URL of its local controlling SIP Proxy Server which was programmed into the RCG during the configuration process. The Main CPU 19 passes this SIP packet off to the Modem/DAA 41 for transmission over the network. The RCG Modem/DAA 41 is connected to and communicating with a modem in a modem bank located in the Inter

Exchange Carrier's (IXCs) facility. This facility may actually be located in the LEC's physical plant or it may be physically located at another location.

[0051] Referring to Figure 2, User A RCG 44 is connected via its modem to the LEC's local Class 5 Office A 47. In this example, the service provider's connections to the IXC A Modem Bank & Router 48 are through IXC A Inter-Office Trunks 46. The service provider may or may not be the IXC. If the service provider does not have its own network, then it will purchase capacity from an IXC and route its traffic to that IXC's facilities. The SIP Invite IP Packet that is sent from User A RCG 44 transits the LEC's Class 5 Office A 47 facilities and is routed to the IXC A Modem Bank/Router 48 via the IXC A Inter-Office Trunks 46. It is the job of the router to take the IP packet and route it on to its intended destination, in this case the appropriate SIP proxy server 51. The router uses the URL for the Softswitch/SIP Proxy Server and Registrar 51 contained in the SIP Invite to know where to route the packet. (Note: The Softswitch, SIP Proxy Server and Registrar are logical devices that may reside in a single computer system, or may be implemented as completely separate devices. In this example, the Softswitch, SIP Proxy Server and Registrar are all shown as a single physical device running these three separate processes.) The IXC A Modem Bank/Router 48 sends the SIP Invite IP packet on to Softswitch/SIP Proxy Server and Registrar 51 via an IP based network route 52.

[0052] The SIP Proxy Server process running on Softswitch/SIP Proxy Server and Registrar 51 performs a look up to determine the termination destination for the call using the dialed digits passed along in the SIP Invite IP packet. In this example, the terminating device for User B RCG 56 is via IXC B's Modem Bank/Router 49 then through Class 5 Office B 50 and over User B POTS line 55 and finally to User B RCG 56. Not covered in detail here is the SIP message handling since it is completely covered in IETF RFC2543. This example shows that both the originating and terminating ends of this call are controlled by the same Softswitch/SIP Proxy Server and Registrar 51 device. Where the originating and terminating ends of a call are controlled by different Softswitch/SIP Proxy Server and Registrar devices, an additional level of SIP messaging is used between the two SIP Proxy Server processes of each of the Softswitch/SIP Proxy Server and Registrar devices to locate and control the terminating location of the call. Again, that information is covered in IETF RFC2543 and is repeated here.

[0053] After the call routing is setup using SIP, and User B answers the call, the Softswitch/SIP Proxy Server and Registrar 51 device sends IP address of the originating User A

RCG 44 to the User B RCG 56 and the IP address of User B RCG 56 to User A RCG 44. Once both RCG's have the other's IP address, they then communicate directly to each other using Real Time Protocol (RTP), and the Softswitch/SIP Proxy Server and Registrar 51 device is no longer part of the connection. This direct communication is shown as IP based RTP Stream 57 between IXC A Modem Bank/Router 48 and IXC B Modem Bank/Router 49.

[0054] On the receiving end of the call, referring back to Figure 2, User B RCG 56 receives the SIP Invite data stream through its Modem/DAA 41 and passes it on to the Main CPU 19. The Main CPU 19 reassembles the SIP IP packet and interrogates the header information to determine which POTS port, POTS 1 Port 30, POTS 2 Port 29 and POTS 3 Port 28 the call is destined for based upon the dialed number contained in the SIP Invite message. The Main CPU 19 commands the DSP Engine 33 to control the SLIC/CODEC Interface 27 to apply ringing voltage to the selected POTS port to ring the phone. A corresponding SIP 180 (ringing) message is sent back to User A RCG 44 in the reverse direction. When the phone attached to the selected POTS port is answered, the off hook condition detected by the SLIC/CODEC Interface 27 is sent through the DSP Engine 33 to the Main CPU 19 that sends the SIP 200 OK message back to User A RCG 44 to let the RCG know that the call was answered by User B RCG 56 and to turn off the audible ringing tone for User B. Once the call is established, User A RCG 44 sends and ACK F12 SIP message to User B RCG 56 indicating that User A RCG 44 understands that the call setup has been completed.

[0055] The two RCGs now exchange SIP messages that include the type of voice encoding that will be used such as G.729, G.723.1, G.711 etc to configure their RTP streams so that they can reliably transfer the user information between them. At this point the call (voice data) is sent between both RCGs and a normal telephone conversation takes place between the user A and user B.

[0056] Call teardown is accomplished by either user hanging up their telephone. The RCG of the terminating end (the end that hangs up the phone first) detects the on-hook condition and sends a SIP Bye message to the other end. The other end upon receiving the SIP Bye message responds with an OK 200 SIP message and disables the appropriate POTS port that the user was using.

RED HAT 1010
RED HAT V. CAS

[0057] The Speakerphone interface can be used for either end of a telephone conversation in exactly the same way as described above except that the corresponding POTS port that the call was intended for is disabled if the call is answered via the speakerphone.

Call Flow 2:

[0058] Calls between an RCG device and a PSTN telephone:

[0059] Refer to figure 3 for the following discussion.

[0060] In the previous call example, the PSTN network had nothing to do since the call was over a previously established modem connection between two RCGs. In the cases where a call originates or terminates to a standard telephone serviced by a Class 5 office, the process is a bit more complicated.

[0061] When a call originates from a user via an RCG, such as User A RCG 44, the call handling practices are exactly the same up to the point that the SIP Proxy Server process in the Softswitch, SIP Proxy Server and Registrar 51 performs a lookup to determine how to route the call. In this case, the SIP Proxy Server process determines that the call terminates at a standard PSTN telephone attached to the Class 5 Office B 50. Since the call originates as a native VoIP service, it must be converted to a standard switched circuit based service to be able to be delivered to User B's telephone 61. The conversion is performed by the IXC's Media Gateway 60 device. In this example, the SIP Proxy Server process routes the call via the IXC's Media Gateway 60 device instead of through the IXC B Modem Bank/Router 49 as before. The SIP Proxy Server process sends a SIP Invite to the IXC's Media Gateway 60 device, as well as other control information, to direct the Media Gateway 60 device to terminate the call and to apply the appropriate voice compression algorithm to the media stream to convert it between switched circuit PCM format and the selected VoIP protocol. At the same time, the Softswitch process of the Softswitch, SIP Proxy Server and Registrar 51 device sends commands to the Media Gateway 60 device to cause it to route the call's PCM data to a specific channel on the IXC B Interoffice Trunk 54. The Softswitch process then generates an SS7 ISUP message that is sent over the PSTN SS7 signaling network to the Class 5 Office B informing it of the incoming call and on which circuit the call has been placed. It is up to the PSTN at this point to ring the User B telephone 61 and process the call between the IXC's Media Gateway and the User B Telephone 61. The Softswitch process handles the signaling conversion between the PSTN's

RED HAT 1010
RED HAT V. CAS

SS7 network and the IXC's SIP messaging, as well as any other proprietary signaling schemes utilized by the IXC to control its network. The standard SS7 call control messages, such as ringing, busy, ring back etc., all have a comparable SIP call flow control messages. The Softswitch simply maps the SS7 messages into SIP messages and visa-versa while routing the messages to the appropriate network equipment.

[0062] As in the case of a call between two RCGs, the RTP media stream (the actual call data) is routed directly be the IXC A Modem Bank/Router 48 and the IXC's Media Gateway 60 via an IP based RTP Stream 57.

[0063] Again, call teardown is accomplished by either user hanging up their telephone. Depending which end hangs up the phone first, the associated network will generate the appropriate messages which will eventually make it back to the Softswitch process. The Softswitch process will relay the appropriate call control messages between the PSTN and the IXC's SIP based control system.

[0064] A telephone call originating from a PSTN telephone is configured and setup in the same fashion as the above description except that the originating messaging begins within the PSTN network instead of the IXC's network.

Personal Computer based communications:

[0065] Where the RCG is connected to a Personal Computer (PC) through either its USB Port 11 or Ethernet Port 12, the RCG operates exactly as a cable or DSL type modem device. The RCG is simply a pass through for the computer's data. The RCG receives the IP packets (of whatever type) from the computer and sends the information unchanged to the Modem/DAA 41 interface for transmission over the Internet. Received information from the Internet is processed in the reverse way.

Dynamic Bandwidth Allocation:

[0066] When a modem connection is established between the RCG and the "Network", the total available bandwidth of this connection is used in a shared/prioritized fashion. Voice services (calls) take precedence over data services. Therefore, when no calls are in progress all of the available bandwidth is used by data type services such as surfing the Internet or uploading/downloading files. However, when a call is to be passed through the RCG, the

required bandwidth for the call is deallocated from the data services and prioritized for the telephone call. This necessary because the Quality Of Service (QOS) is important for effective communications over electronic networks. Since the voice data is carried over a "data centric" network topology (an IP based network), call quality can suffer greatly. Every effort must be taken to eliminate or reduce the packet delay through every network element to be able to provide the best call quality. For data applications such as surfing the Internet or downloading files packet delay is acceptable and expected, and it does not affect the "quality" of the data.

[0067] Since all voice and data is transmitted as packets, there are no "time slots" or other type of "channelization" associated with the transmission of this data over either the POTS line or the wireless interface. Channelization of the voice data does occur within the SLIC/Codec Interface – 27 but is converted into IP packets by the DSP engine – 33 and is handled as packets from that point on.

[0068] Regardless of whether the call is originating or terminating at the RCG, the bandwidth allocation process is the same. When the SIP call setup messages are processed, the type of voice compression that will be used is defined. The RCG Main CPU 19 handles the IP packet flow between the DSP Engine 33, USB Port 12, Ethernet Port 11, 802.11b/g Module 36 and the Modem/DAA 41 devices giving priority to the voice packets to and from the DSP Engine 33 and the Modem/DAA 41 interfaces. Voice compression algorithms create IP packets at a specified number of packets per second. These packets are handled by a timed process in the RCG. This process is started during call setup so that the packets are handled at the correct rate based upon the compression algorithm's packet generation rate. There is no priority assigned between the voice packets of each of the three different POTS Ports 28-30. When more than one voice call is in progress, they are handled in a round robin fashion based upon the packet rate for the compression algorithm used. If, for example, two calls are using a compression scheme that generates packets every 2 milliseconds, and the third generates packets every 4 milliseconds, then the calls utilizing the faster packet rate will be polled at twice the rate as the call using a slower packet rate but they are not given priority "over" the other call. Only when there are no voice packets ready for transmission to the Modem/DAA 41, i.e. during the approximately 2 to 4 millisecond packetization period between packets, will any pending data packets received from the USB Port 12, Ethernet Port 11 or the 802.11b/g Module 36 be sent to the Modem/DAA 41 for transmission over the network.

[0069] Thus, the POTS bandwidth is allocated on a per packet basis, not on a "time slot" or "channelization" or even a "frequency" basis as it the case with T-Carrier, DSL or channelized DSL systems. The same is true for the wireless interface, whether it is carrying pure data or voice.

[0070] Security is currently a big concern in today's networks, and especially with wireless 802.11 transmissions. For this reason, the RCG defaults to route voice data over the users POTS connection via Incoming POTS Port 40. As security concerns regarding wireless communications transmissions are addressed and alleviated, voice call data can, and will, be routed over the wireless interface via the 802.11 b/g Module 36.

FAX Transmission:

[0071] Group 1, 2 and 3 fax transmission is completely supported by the RCG. Fax transmission has the same priority as other voice communications. That is to say that FAX transmission is treated the same as a voice call and takes precedence over native IP data traffic.

Broadband Over POTS:

[0072] One of the primary functions of the RCG is to be able to transfer large files such as video files over the POTS lines to a residence. The ability to do this is limited by the maximum of 56Kbps due to the design of the digital Class 5 office linecards and network transmission schemes based upon T carriers. Typical Class 5 linecard design samples the POTS lines at a rate of 8,000 samples per second. Typically, each sample is digitized into 12 to 16 bits by a codec device. The digitized sample is then companded to an 8 bit sample (1 byte) which makes up the PCM data stream that is transferred over the PSTN. Therefore, each POTS line is sampled at an effective rate of 64Kbps. Since legacy T1 transmission equipment uses robbed-bit signaling, the maximum guaranteed transmission rate for the PSTN as a whole is 56Kbps. Because of this limitation, standard POTS lines can only guarantee a maximum of 56Kbps transfer rate under the best conditions. DSL equipment gets around this problem by installing DSL Access Multiplexing (DSLAM) equipment at the Class 5 office end of the POTS line. By routing around the linecard, as well as routing the data directly over the Internet or other private IP based networks, and therefore the 64Kbps sampling rate, DSL is not bandwidth limited like PSTN traffic is. DSL, however, suffers from the very high costs associated with the purchase,

installation, maintenance, floor lease costs, etc. DSL also suffers from very high deployment costs resulting from the labor intensive process to turn up each line. Each DSL "POTS" line must be inspected, conditioned and all load coils removed.

[0073] The RCG gets around the 56Kbps POTS limitation as well as the DSL problems by using the standard POTS lines as they are, and not requiring any additional equipment to be installed at the Class 5 end of the POTS line. This technique simplifies deployment and drastically cuts costs.

Multilink PPP connections:

[0074] Referring to figure 7, the RCG implements broadband over POTS by employing an 802.11 b/g wireless interface to configure a multilink PPP bundle – 87 (per RFC 1990) comprised of the POTS lines connected to up to 32 RCGs. The limitation of 32 RCGs in a multilink PPP bundle is an arbitrary number and can be scaled up or down as required to support changes in network equipment capabilities. An 802.11b/g interface – 87 is shown as the preferred embodiment, since its 11/54 Mbps interface is considered sufficient bandwidth at this time also, its transmission distances are sufficient for implementation of this device, as opposed to other shorter ranged wireless transmission schemes. However as time goes on, some other wireless interface may be more desirable. There is nothing special or specific to the 802.11 wireless interface that makes it particularly suited to this application. Any wireless interface scheme can be used to implement this device.

[0075] The 802.11b/g interface – 88 on an individual RCG is also used to provide a peer-to-peer wireless home networking environment to other 802.11 compatible wireless component within the residence (such as a hand-held computer device) while acting as an always on dialup Internet gateway over the POTS line – 40 for the other wireless devices. An RCG that is transferring a very large file can request the services of 31 (in this embodiment) other RCGs to assist in the file transfer using their respective POTS lines. A single RCG can therefore gain the combined bandwidth of up to 32 POTS lines transferring data at as much as 56Kbps each to transfer information. This is roughly equivalent to a transfer rate of 1.79Mpbs in the inbound direction (towards the RCG) and approximately 1.075Mbps in the outbound direction (towards the PSTN).

RED HAT 1010
RED HAT V. CAS

[0076] The multilink PPP connection – 87 is initiated by an RCG when it determines that the file (or streaming service) that it is to download (receive) is of sufficient size or requires a minimum amount of bandwidth to justify the use of a multilink PPP connection. This determination can be made on many different criteria such as the file size, type of data being transmitted, the type of compression if any, a Quality-of-Service requirement or any other pertinent information. (Since a certain amount of overhead is required to set up a Multilink PPP connection, it doesn't make sense to use it for small or low data rate transfers.)

Multilink PPP bandwidth allocation and initialization:

[0077] Referring to figure 7, an RCG – 85 that decides to initiate a multilink PPP session with a remote computer i.e. internet URL, requests a multilink PPP session with that URL resource via Link Control Protocol (LCP) session management. The initiating RCG - 85 and the remote server computer (not shown) will negotiate the pertinent multilink parameters such as compression, packet size, sync/async, link quality monitoring, bit stuffing etc. As part of a successful multilink PPP session setup, the initiating RCG –85 receives a multilink PPP session ID, or "magic number", which is used to identify this particular multilink PPP session and is also used for adding or removing links from the multilink bundle – 87 . Once the multilink bundle is setup, the initiating RCG – 85 begins data transfer across its POTS interface – 90 with the far end computer system. The initiating RCG – 85 then looks up the addresses of up to 31 (in this embodiment) other RCGs in its router tables to use for the multilink PPP transfer. (If a Neighborhood Access Point is available the RCG could request service through it instead of setting up a Multilink PPP connection – see figure 8.) The initiating RCG - 85 determines the number of remote RCGs - 86 that it wants to use to get the desired bandwidth for the transfer and sends multilink PPP service requests via the 802.11 b/g wireless connections – 88 to the selected remote RCGs - 86. The selected remote RCGs - 86 respond back to the initiating RCG with either an acceptance or denial of multilink PPP service request. This decision is made based upon the remote RCG's currently available POTS bandwidth and the current local demands for that bandwidth. The remote RCG – 86 makes a decision to participate in a multilink PPP service for an initiating RCG if its POTS line – 91 has a current average bandwidth utilization of less than some arbitrary limit such as 8Kbps to 16 Kbps. This 8Kbps to 16Kbps bandwidth utilization represents about two concurrent voice conversations using standard voice

compression techniques, or some small amount of data transmission occurring on the remote RCG's POTS interface – 91 or a combination thereof. Assuming that the remote RCG – 86 has a modem connection speed somewhere between 33Kbps to 56Kbps, this would leave approximately 17Kbps to 40Kbps of POTS bandwidth that the remote RCG – 86 could offer to the initiating RCG – 85 for use in its multilink PPP bundle - 87. Of course, if there is no activity on the remote RCG – 86 POTS line, then the entire bandwidth will be used for the multilink transfer. If the remote RCG – 86 accepts the multilink PPP service request from the initiating RCG -85 it also sends its currently available bandwidth on its POTS line - 91 that the initiating RCG –85 will have available to it for the multilink PPP transfers. The initiating RCG then selects which of the responding remote RCGs - 86 it will use for the multilink PPP bundle - 87 and sends them a multilink initialization packet containing the URL of the destination as well as the multilink PPP session ID (magic number) and any other pertinent information required for the remote RCG to join in the existing multilink PPP bundle - 87. Each remote RCG POTS interface – 91 participating in a multilink PPP bundle, represents a logical "link" in the initiating RCG's multilink PPP bundle - 87. The remote RCGs – 86 then send a multilink PPP packet (over their respective POTS lines - 91) to the destination URL with the multilink PPP session ID and request to be added to that multilink PPP bundle - 87. The destination URL accepts the request of each of the remote RCGs - 86 and begins sending data packets to each over their respective POTS lines - 91. When the remote RCGs –86 receive data packets sent over their respective POTS lines – 91, that are destined for the multilink PPP bundle - 87, they simply relay the data packets over their 802.11b/g interface - 88 to the initiating RCG -85. It is the responsibility of the initiating RCG – 85 to perform all housekeeping tasks associated with the multilink transfer such as detection of lost packets and data packet reassembly, if those functions are required. Typically, lost packet detection and retransmission and reassembly are performed at higher OSI layers by the requesting software such as a browser. The remote RCGs – 86 are simply data transport nodes and do not perform any data processing on data received over a multilink PPP bundle destined for another RCG.

[0078] The initiating RCG- 85 of the multilink PPP bundle -87 continually monitors the bandwidth of each remote RCG's -86 multilink link which together comprise its multilink bundle – 87. When the available bandwidth for the multilink PPP link of an individual remote RCG -86 falls below an arbitrary minimum set by the initiating RCG – 85, the initiating RCG – 85

RED HAT 1010
RED HAT V. CAS

instructs the offending remote RCG, via its 802.11b/g wireless interface – 88, to remove itself from the multilink bundle. The remote RCG - 86 will then notify the remote computer (URL) to be removed as a link in the current multilink bundle on which it is participating. Alternatively, a remote RCG - 86 can inform the initiating RCG – 85 of local demands on its POTS bandwidth, and request that the initiating RCG – 85 remove it (the remote RCG – 86) from the participation in the initiating RCG's multilink bundle – 87. The remote RCG – 86 will not unilaterally remove itself from participation in a multilink bundle without prior notification to the Initiating RCG - 85. The remote RCG – 86 must get authorization from the initiating RCG – 85 to remove itself from the bundle, since the initiating RCG – 85 will have no way of knowing if the loss of data from that remote RCG is due to some sort of failure or from the remote RCG simply removing itself from participation in the multilink bundle - 87. Also, for services that require a minimum Quality of Service (QOS) such as streaming video, the initiating RCG – 85 needs to control the multilink bundle – 87 bandwidth, and it can not do that if remote RCGs - 86 drop out of the multilink bundle – 87 at will. When a remote RCG – 86 requests to be removed from participation in a multilink bundle - 87, the initiating RCG – 85 must immediately grant the request. If the initiating RCG – 85 were to not grant the request, then the remote RCG – 86 could, in worst case, delay or drop incoming multilink packets destined for the initiating RCG – 85. If this were to happen, the initiating RCG – 85 will have missing packets and could over run its reassembly buffer (if reassembly is being performed by the RCG) causing catastrophic data loss and potentially have to restart the entire transmission. In the least case, the initiating RCG – 85 would have to initiate retransmission of potentially many packets which it assumes to be lost, but may simply be delayed by the remote RCG – 86. Once a remote RCG – 86 has been removed from participating in a multilink bundle – 87, the initiating RCG – 85 will then look up another remote RCG in its router tables that can be used to participate in the multilink PPP bundle – 87, and request that the new remote RCG -86 to join the existing multilink bundle – 87. This is necessary to maintain a minimum Quality of Service. Since RCGs prioritize their bandwidth requirements based upon their own "local" service requirements, their participation in any multilink PPP service is available only to bandwidth that is not required for its own current local demands. Therefore, a remote RCG's – 86 bandwidth that is available for multilink services can change at any time. The dynamic reassignment of remote RCGs that participate in a multilink bundle is also required to prevent a remote RCG – 86 from receiving packets so slowly

RED HAT 1010
RED HAT V. CAS

that the originating RCG – 85 interprets this delay as lost packets. When the initiating RCG – 85 determines that a packet has been lost, and it is performing packet reassembly, it will, itself, request that the remote computer retransmit the lost packet. The initiating RCG will not ask any remote RCG – 86 to perform retransmissions or any other kind of data computations or manipulations.

[0079] Refer to figure 8 for the following discussion. In areas where there are existing 802.11 network equipment such as Neighborhood Access Point - 93, or where there is sufficient bandwidth requirements from an installed base of RCGs to warrant the deployment of a Neighborhood Access Point - 93, the RCGs will automatically detect the presence of the compatible neighborhood access points and register with that equipment for transport services. In this case, an RCG has an alternative broadband service route in addition to the multilink option. An RCG can communicate directly with a local 802.11 Network Access Point - 93 that is connected to the PSTN or another carrier's network via some sort of high bandwidth backbone such as a T-Carrier or optical carrier connection (T1, T3, OC3, OC12 etc.). The RCGs can communicate directly with neighborhood access points at their full 11/54 Mbps data rates to provide up to that full bandwidth to the user. This functionality is, of course, subject to the Network Access Point – 93 allowing access to its resources to the RCG according to the 802.11 wireless protocol.

[0080] This architecture is important since it provides an avenue to relieve the local class 5 offices of congestion that can be caused by a wide spread deployment of RCGs within the class 5 service area. Since each RCG requires a constant 24/7 modem connection, this ties up the class 5 switch in a fashion that it was not designed for. This could lead to wholesale class 5 office expansions which would be extremely expensive, and potentially prohibitive for the Local Exchange Carriers (LECs). By deploying neighborhood access points that operate at 54Mbps (802.11a/g), in areas where the RCG concentration starts placing too high of a demand on the local class 5 office, the bandwidth from the RCGs can be aggregated by the neighborhood access points into large capacity backbones to relieve the switch congestion. In this case, the RCG POTS interface – 40 can be used solely as a lifeline connection, or when the 802.11 wireless LAN in which the RCG resides is congested. With this implementation, voice calls will be carried over the wireless interface as opposed to the POTS interface on which they would normally be carried. Also, the RCG modem - 41 connected to the POTS interface – 40 would

RED HAT 1010
RED HAT V. CAS

not need to maintain a constant always on connection, it could be a dialed on demand connection and thereby relieving the local class 5 office switch.

Bouncing or Hopping traffic:

[0081] Refer to figure 9. In areas where the density of RCGs is less than the arbitrary number of 32 RCGs within its transmission radius for multilink connections, the RCGs can bounce, or relay, the data traffic between each other to reach other RCGs located farther than the RCGs wireless transmission area. Since the POTS data rate of 56Kbps is miniscule as compared to the 11/54 Mbps 802.11b/g transfer rate, the overhead associated with bouncing data traffic is not a concern. This will allow for the fastest possible data transfer across POTS lines for every RCG user.

[0082] RCGs can also communicate directly with an 802.11a Neighborhood Access Point - 93 device (such as that used in WI-FI networks) that is connected to a data network connection such as a DS3 or OC3 service via a Metro Access Point. Where an RCG has access to a broadband device such as an 802.11a type device, either directly or via a bounced signal across several RCGs, a direct high speed data connection can be used to transfer large files. When an RCG is communicating directly with an 802.11a device, data transfer speeds up to the full 11/54 Mbps can be achieved to speed video transfers for such applications as video on demand.

RCG Router Tables:

[0083] The RCG continually monitors its 802.11b/g wireless interface for transmissions from other RCGs or 802.11a Access Points as well as other 802.11b/g wireless devices within the residence. The RCG provides a means to identify and store the addresses of other RCGs and Neighborhood Access Points within its local area. The RCG creates routing tables, much like an IP router device, so that it knows which RCGs are local to itself, and how many hops will be needed to reach the maximum of 32 POTS lines or a Neighborhood Access Point. The RCG can then determine how to route its data transfers to gain the most efficient routing and network efficiencies.

[0084] The router tables in each RCG are continually updated by monitoring the 802.11 frames received from other RCGs or Network Access Points. The number of entries in the router tables can be limited to an arbitrary number of other RCGs but contains room for at least 32

entries. For each RCG listed in the router tables, certain information is kept such as signal strength, if the other device is another RCG or some other 802.11 compatible device, if the other RCG is reachable directly, or if it is accessed via another RCGs by hopping. In the case there RCG hopping is used, the address of each RCG that is hopped is stored in the routing table. When an RCG is first powered up, it monitors its 802.11 interface for a period of time to identify other RCGs or Neighborhood Access Point devices, or other Wi-Fi devices within its transmission area. When first deployed, the RCG will not have any entries in its routing tables. This condition can persist for some time, until another RCG or other 802.11 device is deployed within the former RCGs transmission range. In this state, the RCG can be used for wireless networking within the residence, connecting the local computer to the RCGs always on Internet connection as well as to other wireless devices within the home. When an 802.11 frame is received by an RCG that contains the address of an RCG whose address does not exist in the receiving RCG's router table (as is the case when a new RCG is first installed), the receiving RCG polls the transmitting RCG for its router table information. The received router information is reviewed for other RCG's or other 802.11 compatible devices that are not listed in its own router tables. A decision is made to use some, all, or none of the new route information in its router tables. There is no real limitation on the number of RCGs that can be "hopped" in order to build a complete router table or to reach many more RCGs, Neighborhood Access Points or Metro Access Points.

Equipment and Services Priorities:

[0085] Just as voice has priority over data transfers within a single RCG, local RCG traffic (both voice and data) has priority over remote transfer requests by other RCGs. If an RCG is currently using all of its available POTS bandwidth for its own purposes, it will deny any remote RCG's request for access to its POTS line. Only when there is available POTS bandwidth available will an RCG allow other RCGs access to its POTS line. It will, however, bounce the traffic to other RCGs within its transmission area if requested by the originating RCG. In this case, available wireless bandwidth of any RCG (that amount of the 11/54 Mbps not currently used for its own traffic) can be used to relay traffic between other RCGs that do have available POTS bandwidth. In the same fashion, excess wireless bandwidth of any RCG can be by other RCGs to access high speed backbones via Neighborhood or Metro Access Points.

[0086] RCGs can also access other 802.11 wireless networks (Wi-Fi for example) that happen to be deployed within their service areas. In this case, access agreements between the RCG service provider and the wireless network operator must be put into place. By remaining compatible with the 802.11 standards, the RCGs capabilities will benefit by the growing WI-FI network infrastructure, an the WI-FI network will benefit by the RCG network.

Voice Over POTS and Data Over Wireless:

[0087] Normally, all communications from an RCG will be transmitted over the local POTS line, or a multilink bundle made up of many remote RCG POTS lines. By default, all voice calls will be carried over the "local" POTS line – 40 to alleviate security/privacy concerns and only "data" will be transmitted via multilink bundles. Generally, since voice "data" is compressed locally by the RCG, the required bandwidth for a voice call is so small that it does not make for a good choice to route it over a multilink connection. However, it is possible to route voice calls entirely over another RCGs POTS line. It is also possible for an RCG to transmit voice data over POTS lines connected to other RCGs via multilink bundles or directly over its 802.11 wireless interface, and since the 802.11 specification does support the Wired Equivalent Privacy (WEP) specification, a secured connection can be made in this fashion.

Remote Upgradability

[0088] The RCG is designed with the capability for unassisted remote upgradability. The remote upgradability feature is not the same as the storing of operational parameters during the configuration process. This feature allows the system operator to add new features as they are developed, or to correct bugs remotely and without the users knowledge or involvement. This feature can also be used to install a complete new operating system (OS).

[0089] Key to the successful implementation of the Remote upgradability feature is that a completely fail-safe routine be implemented that can recover from any failure during the upgrade process and to re-initiate and monitor the upgrade process until it has successfully completed. This process is carried out by loading a boot loader program that is stored in non-volatile memory. The boot loader program provides the minimum capabilities to start up the RCG and connect into the system operator's diagnostics network to download fixes or a new OS. The boot loader program is technically part of the OS, but it cannot be over written by any upgrade

process. It is responsible for detecting failure of the RCG and initiating a connection to the diagnostics network. Once connected with the diagnostics network, the boot loader will provide diagnostics support to help correct the RCG's operation and reconfigure the OS.

CLAIMS

What is claimed is:

1. A method for establishing and sending data across multiple paths between a server and an initiating device, comprising:

receiving at the server information of each of a plurality of devices selected to participate in a multipath connection, the multipath connection comprising each the devices being connected with the initiating device; and

the server performing at least one of (i) dividing data into groups of data packets and sending different ones of the groups of data packets to each of the devices participating in the multipath connection, (ii) sending separate streams of data to each of the devices participating in the multipath connection, and (iii) sending combinations of separate streams and groups of data packets to each of the devices participating in the multipath connection.

2. The method of claim 1, wherein the information includes a connection ID that is is generated by one of the server, one of the plurality of devices, the initiating device and a network device.

3. The method of claim 2, further comprising receiving at the server information of additional devices able to participate in the multipath connection, the additional devices being associated with the connection ID.

4. The method of claim 2, further comprising receiving at the server information about which devices are to be removed from the multipath connection, the devices to be removed being associated with the connection ID.

5. The method of claim 1, further comprising the server determining which devices are to be removed from the multipath connection.

6. The method of claim 1, wherein the server is one of a multipath link aggregator, network relay, and multipath link server.

7. A method for establishing and sending data across multiple paths in different multipath connections from a server, comprising:

the server receiving connection requests for a plurality of multipath connections;

the server associating a connection ID for each connection request associated with each of the multipath connections;

the server receiving information of each of a plurality of devices participating in a particular multipath connection having a particular connection ID; and

the server sending data associated with the particular connection ID to each of the devices participating in the multipath connection associated with the connection ID.

8. The method of claim 7, further comprising receiving at the server information of additional devices participating in the particular multipath connection, the additional devices being associated with the particular connection ID.

9. The method of claim 7, further comprising receiving at the server identification of devices to be removed from the particular multipath connection, the devices to be removed being associated with the particular connection ID.

10. The method of claim 7, further comprising the server determining which devices are to be removed from the multipath connection.

11. The method of claim 7, wherein the server is one of a multipath link aggregator, network relay, and multipath link server.

12. The method of claim 7, wherein the connection ID is generated by one of the server, one of the plurality of devices, or a network device.

13. A method for establishing and sending data across multiple paths between a device and a server, comprising:

selecting one or more remote devices to participate in a multipath connection, wherein the multipath connection comprises a connection between the device and each of the selected one or more remote devices;

receiving data concurrently at the device via the selected one or more remote devices during the multipath connection.

14. The method of claim 13, further comprising receiving at the device information about the one or more remote devices confirmation of a capability of participating in the multipath connection.

15. The method of claim 13, wherein the device is one of a cell phone, a tablet, a router, a computer, a Wi-Fi device, a Wi-Max device, a communications device, a mobile communications device, and a Bluetooth device.

16. The method of claim 13, wherein the devices are one or more routers, modems, cell phones, tablets, computers, cellular towers, wireless towers, base stations, wireless base stations, repeaters, wireless repeaters, Wi-Fi devices, networking devices, hot spot devices, Wi-MAX devices, LTE devices, GSM devices, consumer electronic devices, and an Internet- ready device.

17. The method of claim 13, wherein the selected one or more remote devices is capable of connecting to another remote device which is capable of participating in the multipath connection, and further comprising relaying data from the another remote device to the device through the selected one or more remote devices.

18. A communications device comprising:

at least one interface capable of connecting concurrently to a plurality of remote devices; and

a processor coupled to the at least one interface that

selects one or more remote devices to participate in a multipath connection, wherein the multipath connection comprises a connection between the device and each of the selected one or more remote devices;

receiving data concurrently at the device via the selected one or more remote devices during the multipath connection.

19. The communications device of claim 19, wherein the device is one of a cell phone, a tablet, a router, a computer, a Wi-Fi device, a Wi-Max device, a mobile communications device, and a Bluetooth device.

20. The communications device of claim 19, wherein each of the remote devices comprises one or more communications devices, cellular devices, tablet devices, router devices, computer devices, satellite devices, Wi-MAX devices, wireless devices, mobile communications devices, Bluetooth devices, and Wi-Fi devices.

MULTIPATH COMMUNICATION DEVICES AND METHODS

ABSTRACT

The Residential Communications Gateway (RCG) is a broadband communications device that combines all voice, data and video communications to and from a typical residence or small business for transmission over a single, or a plurality of Plain Old Telephone Service (POTS) lines separately or in conjunction with, a wireless broadband backbone. The RCG does this by employing packetized data with Voice over Internet Protocol (VoIP) technologies combined with RF communications technologies. A key consideration to the design of the RCG is that no additional or special transmission equipment must be installed at the Central Office or anywhere else in the network to enable new calling features provided by the RCG as is the case with DSL and Cable systems. By eliminating the requirement for costly infrastructure enhancements, ubiquitous high speed communications and services can be deployed to every POTS subscriber.

16 – CPU USB INTERFACE
12 – USB PORT
11 – ETHERNET PORT
13 – USB DRIVER
21 – DISPLAY AND KEY DRIVERS
10 – DISPLAY AND KEYPAD
15 – USB VIDEO INPUT
17 –ETHERNET DRIVER
ETHERNET DRIVER
USB DRIVER
VIDEO CAMERA
14 – VIDEO CAMERA
DISPLAY AND KEY PAD
22 – DISPLAY AND KEYPAD CABLES
18 – CPU E-NET I/F
42 – CPU TO MODEM/DAA I/F
41 – MODEM/DAA
22 – SPEAKERPHONE DRIVERS
23 – MICROPHONE CABLE
19 – MAIN CPU
40 – INCOMING POTS PORT
V.90/V.92 MODEM/DAA
20 – CPU DISPLAY & KEYPAD I/F
MAIN SYSTEM PROCESSOR
DISPLAY & KEY PAD I/F
37 – LIFELINE BYPASS ROUTING
43 – CPU LIFELINE INT
24 – MICROPHONE
32 – DSP SPKR PHONE I/F
LIFELINE FAIL-SAFE LOGIC
SPEAKERPHON E I/F
25 – SPEAKER
33 – DSP ENGINE
39 – LIFELINE FAILSAFE LOGIC
POWER CIRCUITRY
28 – POTS 3 PORT
CODECS AND SLICS
26 – SPKR CABLE
VOICE PROC-DSP AND IP PACKETIZATION
38 – POWER CONDITIONING AND MONITORING
29 – POTS 2 PORT
3 TELEPHONE INTERFACES
35 – 802.11B I/F
36 – 802.11B/G MODULE
30 – POTS 1 PORT
IRDA I/F
WIFI MODULE
27 – LIFELINE BYPASS ROUTING
27 – SLIC/CODEC INTERFACE
31 – DSP AUDIO INTERFACE
34 – CPU/DSP COMM PATH
8 – IRDA PORT
9 – 802.11B INTERFACE
10 – IRDA Interface

FIG. 1

58 – PSTN NETWORK
59 – IXC NETWORK
SOFTSWITCH, SIP PROXY SERVER AND REGISTRAR
51 – SOFTSWITCH, SIP PROXY SERVER AND REGISTRAR
52 – IP BASED NETWORK ROUTE
53 – IP BASED NETWORK ROUTE
49 – IXC B MODEM/ROUTER
46 – IXC A INTEROFFICE TRUNK
IXC A MODEM BANK/ROUTE
IXC B MODEM BANK/ROUT
54 – IXC B INTEROFFICE TRUNK
45 – USER A POTS LINE
57 – IP BASED RTP STREAM
55 – USER B POTS LINE
48 – IXC A MODEM/ROUTER
RCG
CLASS 5 OFFICE A
CLASS 5 OFFICE B
RCG
50 – CLASS 5 OFFICE B
44 – USER A RCG
47 – CLASS 5 OFFICE A
56 – USER B RCG

FIG. 2

58 – PSTN NETWORK
59 – IXC NETWORK
51 – SOFTSWITCH, SIP PROXY SERVER AND REGISTRAR
52 – IP BASED NETWORK ROUTE
SOFTSWITCH, SIP PROXY SERVER AND REGISTRAR
53 – IP BASED MEDIA GATEWAY CONTROL
46 – IXC A INTEROFFICE TRUNK
60 – IXC MEDIA GATEWAY
54 – IXC B INTEROFFICE TRUNK
45 – USER A POTS LINE
IXC A MODEM BANK/ROUTER
57 – IP BASED RTP STREAM
MEDIA GATEWAY - SIP BASED
55 – USER B POTS LINE
48 – IXC A MODEM/ROUTER
RCG
CLASS 5 OFFICE A
CLASS 5 OFFICE B
50 – CLASS 5 OFFICE B
44 – USER A RCG
47 – CLASS 5 OFFICE A
61 – USER B TELEPHONE

FIG. 3

72 - DISPLAY
Display (optionally a touch screen)
71-75 - PROGRAMMABLE FUNCTION KEYS
77-79 - LINE SELECT BUTTONS
76 - NUMBER KEYPAD
70 - RCG BASE
25 - SPEAKER
1
2
3
4
5
6
7
8
9
*
0
#
46 - WEIRLESS KEYBOARD PORT - IRDA
24 - MICROPHONE
80-82 - LINE ACTIVE LED'S

FIG. 4

Display
72 - DISPLAY
70 - RFD BASE
40 - INCOMING POTS PORT
84 - HEADSET/HANDSET PORT
11 - ETHERNET PORT
12 - USB PORT
28 - POTS 2 PORT
29 - POTS 2 PORT
USB - POTS 1 PORT

FIG. 5

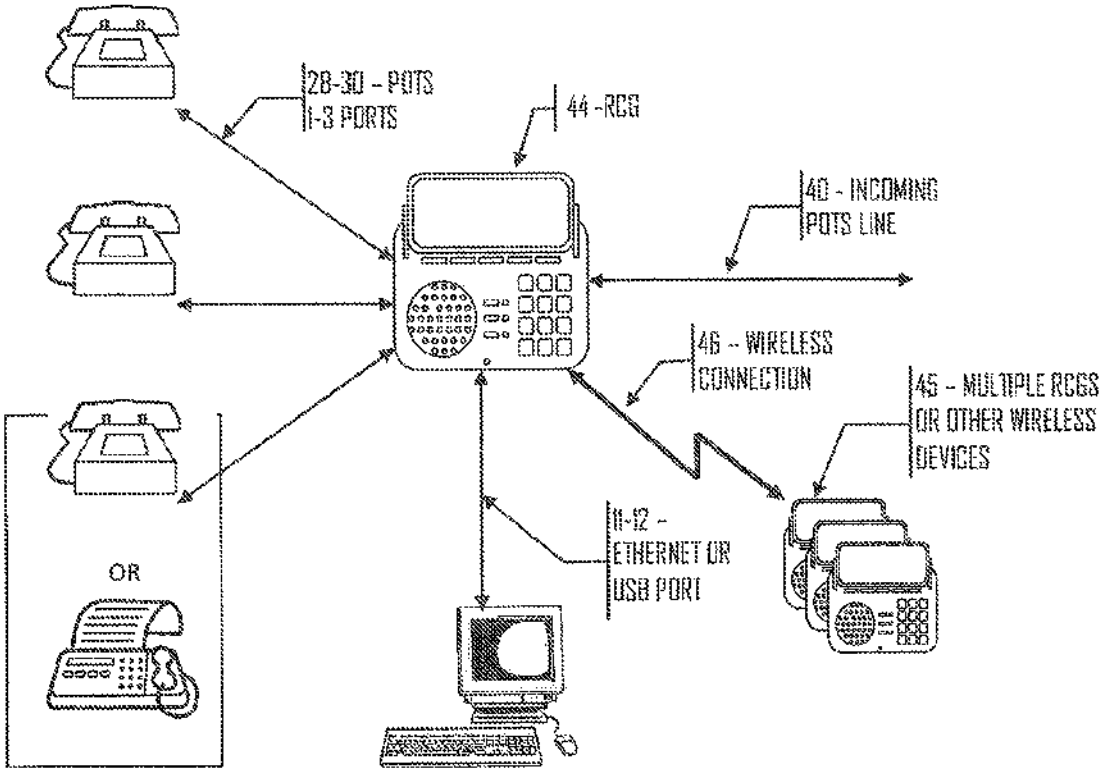


FIG. 6

RED HAT 1010
RED HAT V. CAS

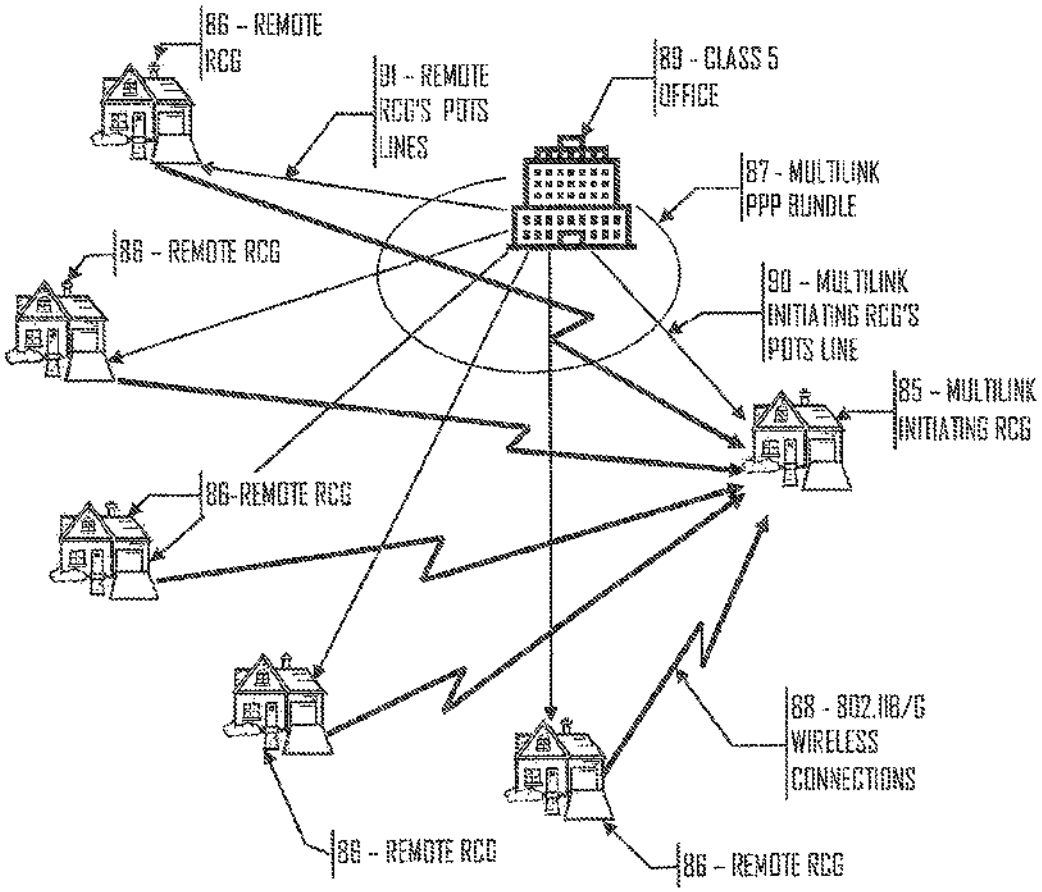


FIG. 7

RED HAT 1010
RED HAT V. CAS

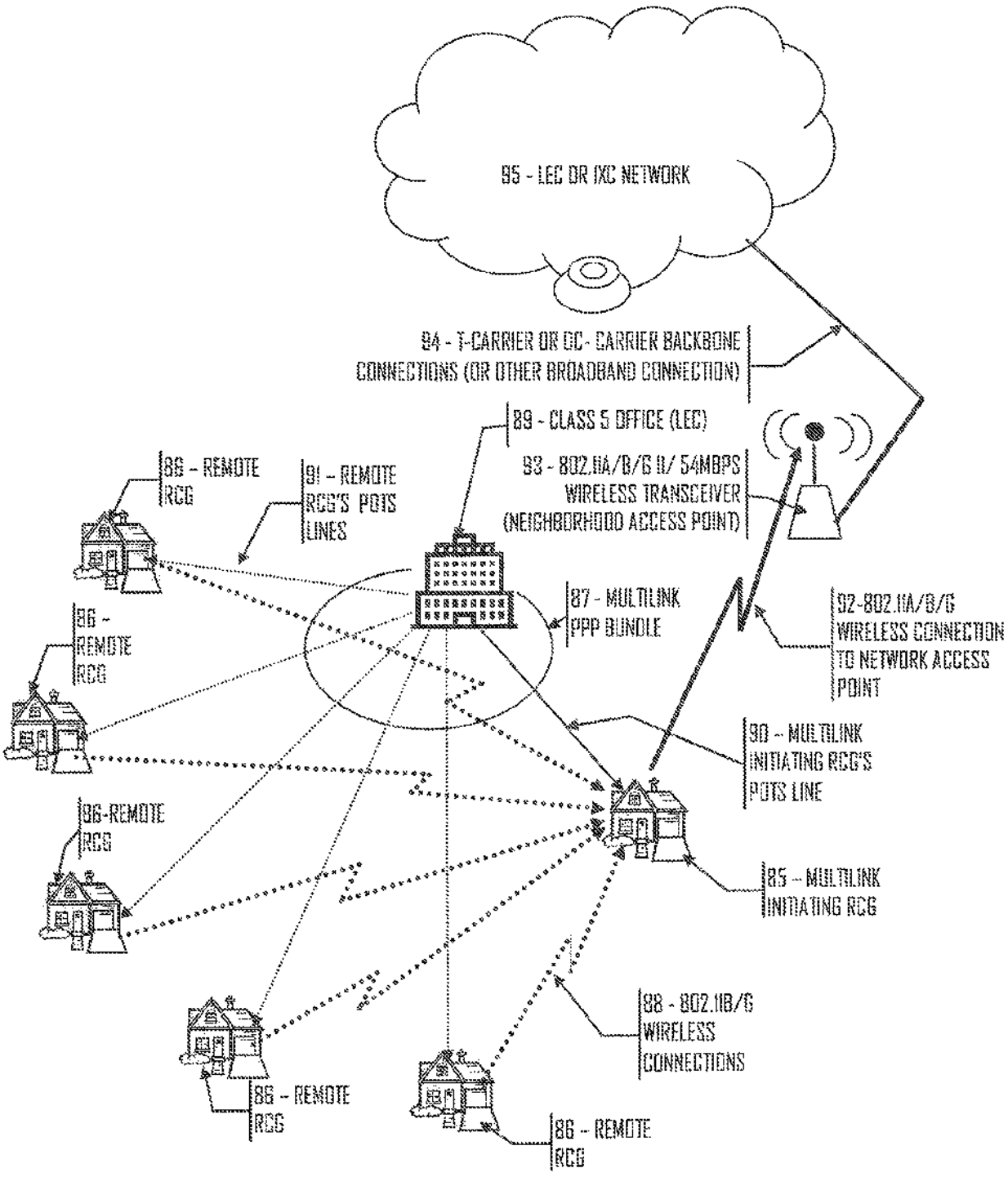


FIG. 8

RED HAT 1010
RED HAT V. CAS

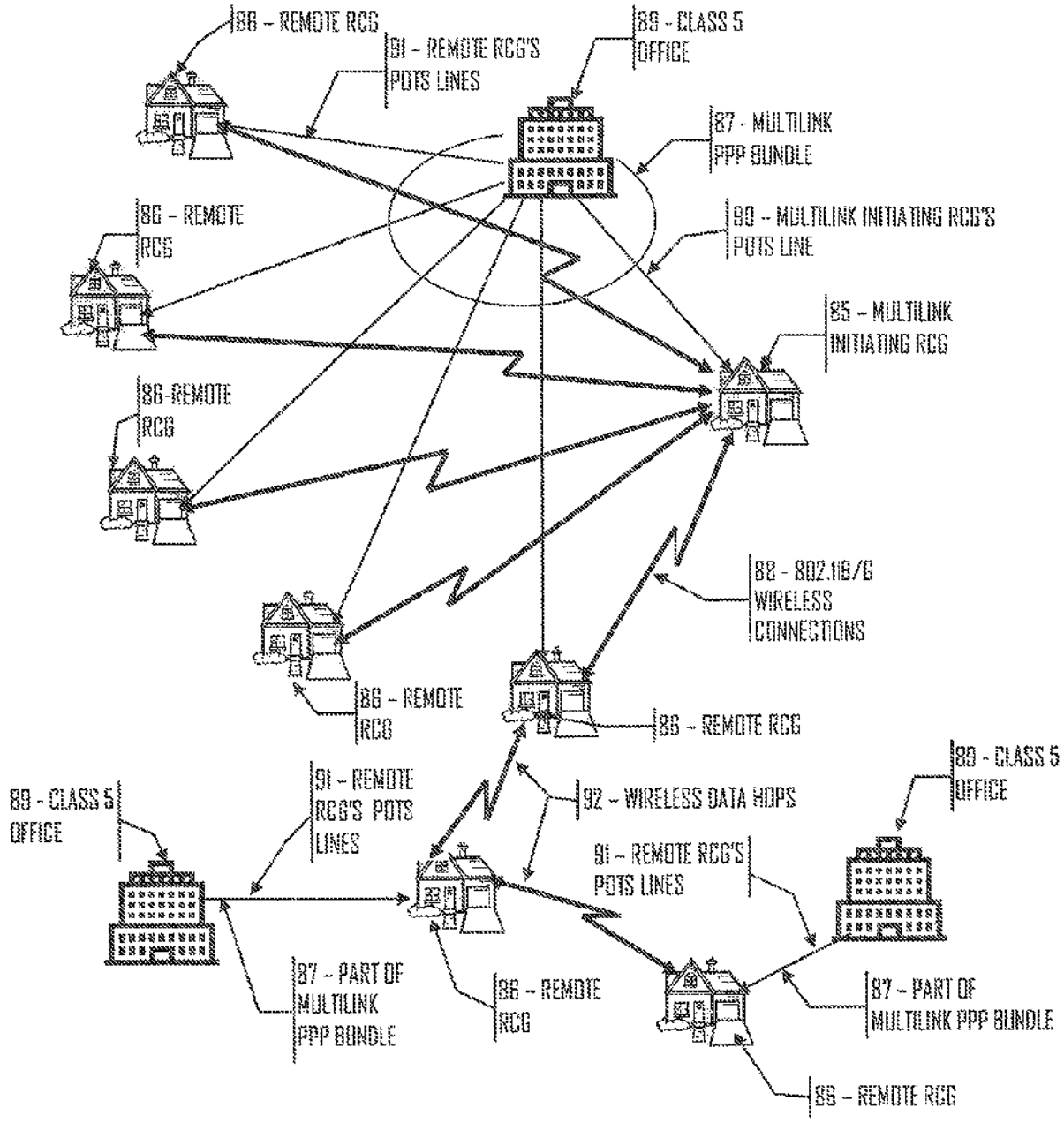


FIG. 9

RED HAT 1010
RED HAT V. CAS

PTO/AIA/01 (06-12)
Approved for use through 01/31/2014. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

DECLARATION (37 CFR 1.63) FOR UTILITY OR DESIGN APPLICATION USING AN APPLICATION DATA SHEET (37 CFR 1.76)

Title of Invention	MULTIPATH COMMUNICATION DEVICES AND METHODS

As the below named inventor, I hereby declare that:

This declaration is directed to:

- [■] The attached application, or
- [ ] United States application or PCT international application number _____

  filed on _____.

The above-identified application was made or authorized to be made by me.

I believe that I am the original inventor or an original joint inventor of a claimed invention in the application.

I hereby acknowledge that any willful false statement made in this declaration is punishable under 18 U.S.C. 1001 by fine or imprisonment of not more than five (5) years, or both.

WARNING:

Petitioner/applicant is cautioned to avoid submitting personal information in documents filed in a patent application that may contribute to identity theft. Personal information such as social security numbers, bank account numbers, or credit card numbers (other than a check or credit card authorization form PTO-2038 submitted for payment purposes) is never required by the USPTO to support a petition or an application. If this type of personal information is included in documents submitted to the USPTO, petitioners/applicants should consider redacting such personal information from the documents before submitting them to the USPTO. Petitioner/applicant is advised that the record of a patent application is available to the public after publication of the application (unless a non-publication request in compliance with 37 CFR 1.213(a) is made in the application) or issuance of a patent. Furthermore, the record from an abandoned application may also be available to the public if the application is referenced in a published application or an issued patent (see 37 CFR 1.14). Checks and credit card authorization forms PTO-2038 submitted for payment purposes are not retained in the application file and therefore are not publicly available.

LEGAL NAME OF INVENTOR

Inventor: ERIC DELANGIS Date (Optional): 10/10/14

Signature: _____

Note: An application data sheet (PTO/SB/14 or equivalent), including naming the entire inventive entity, must accompany this form or must have been previously filed. Use an additional PTO/AIA/01 form for each additional inventor.

This collection of information is required by 35 U.S.C. 115 and 37 CFR 1.63. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 1 minute to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

RED HAT 1010
RED HAT V. CAS

PATENT APPLICATION FEE DETERMINATION RECORD

Substitute for Form PTO-875

Application or Docket Number: 14/512,414

APPLICATION AS FILED - PART I

FOR	(Column 1) NUMBER FILED	(Column 2) NUMBER EXTRA	SMALL ENTITY RATE($)	SMALL ENTITY FEE($)	OR	OTHER THAN SMALL ENTITY RATE($)	OTHER THAN SMALL ENTITY FEE($)
BASIC FEE (37 CFR 1.16(a), (b), or (c))	N/A	N/A	N/A	70		N/A	
SEARCH FEE (37 CFR 1.16(k), (l), or (m))	N/A	N/A	N/A	300		N/A	
EXAMINATION FEE (37 CFR 1.16(o), (p), or (q))	N/A	N/A	N/A	360		N/A	
TOTAL CLAIMS (37 CFR 1.16(i))	20 minus 20=	*	x 40 =	0.00	OR		
INDEPENDENT CLAIMS (37 CFR 1.16(h))	4 minus 3 =	1	x 210 =	210			
APPLICATION SIZE FEE (37 CFR 1.16(s))	If the specification and drawings exceed 100 sheets of paper, the application size fee due is $310 ($155 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).			0.00			
MULTIPLE DEPENDENT CLAIM PRESENT (37 CFR 1.16(j))				0.00			
* If the difference in column 1 is less than zero, enter "0" in column 2.			TOTAL	940		TOTAL	

APPLICATION AS AMENDED - PART II

		(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	SMALL ENTITY RATE($)	SMALL ENTITY ADDITIONAL FEE($)	OR	OTHER THAN SMALL ENTITY RATE($)	OTHER THAN SMALL ENTITY ADDITIONAL FEE($)
AMENDMENT A	Total (37 CFR 1.16(i))	*	Minus	**	=	x =		OR	x =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=	x =		OR	x =	
	Application Size Fee (37 CFR 1.16(s))									
	FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							OR		
						TOTAL ADD'L FEE		OR	TOTAL ADD'L FEE	

		(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	SMALL ENTITY RATE($)	SMALL ENTITY ADDITIONAL FEE($)	OR	OTHER THAN SMALL ENTITY RATE($)	OTHER THAN SMALL ENTITY ADDITIONAL FEE($)
AMENDMENT B	Total (37 CFR 1.16(i))	*	Minus	**	=	x =		OR	x =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=	x =		OR	x =	
	Application Size Fee (37 CFR 1.16(s))									
	FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							OR		
						TOTAL ADD'L FEE		OR	TOTAL ADD'L FEE	

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.

** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20".

*** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3".

The "Highest Number Previously Paid For" (Total or Independent) is the highest found in the appropriate box in column 1.

RED HAT 1010
RED HAT V. CAS

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING or 371(c) DATE	GRP ART UNIT	FIL FEE REC'D	ATTY.DOCKET.NO	TOT CLAIMS	IND CLAIMS
14/512,414	10/11/2014	2652	940	ED 2000	20	4

CONFIRMATION NO. 9530

64828
Harold McGurk
The Law Office of Clay McGurk
P.O. BOX 1488
Orange, CA 92856

FILING RECEIPT

OC000000071390391

Date Mailed: 10/20/2014

Receipt is acknowledged of this non-provisional patent application. The application will be taken up for examination in due course. Applicant will be notified as to the results of the examination. Any correspondence concerning the application must include the following identification information: the U.S. APPLICATION NUMBER, FILING DATE, NAME OF APPLICANT, and TITLE OF INVENTION. Fees transmitted by check or draft are subject to collection. Please verify the accuracy of the data presented on this receipt. **If an error is noted on this Filing Receipt, please submit a written request for a Filing Receipt Correction. Please provide a copy of this Filing Receipt with the changes noted thereon. If you received a "Notice to File Missing Parts" for this application, please submit any corrections to this Filing Receipt with your reply to the Notice. When the USPTO processes the reply to the Notice, the USPTO will generate another Filing Receipt incorporating the requested corrections**

Inventor(s)

Eric M. Delangis, Huntsville, AL;

Applicant(s)

Eric M. Delangis, Huntsville, AL;

Power of Attorney: The patent practitioners associated with Customer Number 64828

Domestic Priority data as claimed by applicant

This application is a CON of 13/531,294 06/22/2012 PAT 8861349 *
which is a CON of 12/581,852 10/19/2009 PAT 8228801
which is a CON of 10/686,375 10/14/2003 PAT 7606156
which claims benefit of 60/418,521 10/15/2002
(*)Data provided by applicant is not consistent with PTO records.

Foreign Applications for which priority is claimed (You may be eligible to benefit from the **Patent Prosecution Highway** program at the USPTO. Please see http://www.uspto.gov for more information.) - None.
Foreign application information must be provided in an Application Data Sheet in order to constitute a claim to foreign priority. See 37 CFR 1.55 and 1.76.

If Required, Foreign Filing License Granted: 10/16/2014
The country code and number of your priority application, to be used for filing abroad under the Paris Convention, is **US 14/512,414**
Projected Publication Date: Perfected
Non-Publication Request: No
Early Publication Request: No
**** SMALL ENTITY ****

Title

MULTIPATH COMMUNICATION DEVICES AND METHODS

Preliminary Class

379

Statement under 37 CFR 1.55 or 1.78 for AIA (First Inventor to File) Transition Applications: No

PROTECTING YOUR INVENTION OUTSIDE THE UNITED STATES

Since the rights granted by a U.S. patent extend only throughout the territory of the United States and have no effect in a foreign country, an inventor who wishes patent protection in another country must apply for a patent in a specific country or in regional patent offices. Applicants may wish to consider the filing of an international application under the Patent Cooperation Treaty (PCT). An international (PCT) application generally has the same effect as a regular national patent application in each PCT-member country. The PCT process **simplifies** the filing of patent applications on the same invention in member countries, but **does not result** in a grant of "an international patent" and does not eliminate the need of applicants to file additional documents and fees in countries where patent protection is desired.

Almost every country has its own patent law, and a person desiring a patent in a particular country must make an application for patent in that country in accordance with its particular laws. Since the laws of many countries differ in various respects from the patent law of the United States, applicants are advised to seek guidance from specific foreign countries to ensure that patent rights are not lost prematurely.

Applicants also are advised that in the case of inventions made in the United States, the Director of the USPTO must issue a license before applicants can apply for a patent in a foreign country. The filing of a U.S. patent application serves as a request for a foreign filing license. The application's filing receipt contains further information and guidance as to the status of applicant's license for foreign filing.

Applicants may wish to consult the USPTO booklet, "General Information Concerning Patents" (specifically, the section entitled "Treaties and Foreign Patents") for more information on timeframes and deadlines for filing foreign patent applications. The guide is available either by contacting the USPTO Contact Center at 800-786-9199, or it can be viewed on the USPTO website at http://www.uspto.gov/web/offices/pac/doc/general/index.html.

For information on preventing theft of your intellectual property (patents, trademarks and copyrights), you may wish to consult the U.S. Government website, http://www.stopfakes.gov. Part of a Department of Commerce initiative, this website includes self-help "toolkits" giving innovators guidance on how to protect intellectual property in specific countries such as China, Korea and Mexico. For questions regarding patent enforcement issues, applicants may call the U.S. Government hotline at 1-866-999-HALT (1-866-999-4258).

LICENSE FOR FOREIGN FILING UNDER

Title 35, United States Code, Section 184

Title 37, Code of Federal Regulations, 5.11 & 5.15

GRANTED

The applicant has been granted a license under 35 U.S.C. 184, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" followed by a date appears on this form. Such licenses are issued in all applications where the conditions for issuance of a license have been met, regardless of whether or not a license may be required as set forth in 37 CFR 5.15. The scope and limitations of this license are set forth in 37 CFR 5.15(a) unless an earlier license has been issued under 37 CFR 5.15(b). The license is subject to revocation upon written notification. The date indicated is the effective date of the license, unless an earlier license of similar scope has been granted under 37 CFR 5.13 or 5.14.

This license is to be retained by the licensee and may be used at any time on or after the effective date thereof unless it is revoked. This license is automatically transferred to any related applications(s) filed under 37 CFR 1.53(d). This license is not retroactive.

The grant of a license does not in any way lessen the responsibility of a licensee for the security of the subject matter as imposed by any Government contract or the provisions of existing laws relating to espionage and the national security or the export of technical data. Licensees should apprise themselves of current regulations especially with respect to certain countries, of other agencies, particularly the Office of Defense Trade Controls, Department of State (with respect to Arms, Munitions and Implements of War (22 CFR 121-128)); the Bureau of Industry and Security, Department of Commerce (15 CFR parts 730-774); the Office of Foreign AssetsControl, Department of Treasury (31 CFR Parts 500+) and the Department of Energy.

NOT GRANTED

No license under 35 U.S.C. 184 has been granted at this time, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" DOES NOT appear on this form. Applicant may still petition for a license under 37 CFR 5.12, if a license is desired before the expiration of 6 months from the filing date of the application. If 6 months has lapsed from the filing date of this application and the licensee has not received any indication of a secrecy order under 35 U.S.C. 181, the licensee may foreign file the application pursuant to 37 CFR 5.15(b).

SelectUSA

The United States represents the largest, most dynamic marketplace in the world and is an unparalleled location for business investment, innovation, and commercialization of new technologies. The U.S. offers tremendous resources and advantages for those who invest and manufacture goods here. Through SelectUSA, our nation works to promote and facilitate business investment. SelectUSA provides information assistance to the international investor community; serves as an ombudsman for existing and potential investors; advocates on behalf of U.S. cities, states, and regions competing for global investment; and counsels U.S. economic development organizations on investment attraction best practices. To learn more about why the United States is the best country in the world to develop technology, manufacture products, deliver services, and grow your business, visit http://www.SelectUSA.gov or call +1-202-482-6800.

To: CLAYMCGURK@GMAIL.COM,,
From: PAIR_eOfficeAction@uspto.gov
Cc: PAIR_eOfficeAction@uspto.gov
Subject: Private PAIR Correspondence Notification for Customer Number 64828

Oct 20, 2014 05:21:42 AM

Dear PAIR Customer:

Harold McGurk
The Law Office of Clay McGurk
P.O. BOX 1488
Orange, CA 92856
UNITED STATES

The following USPTO patent application(s) associated with your Customer Number, 64828 , have new outgoing correspondence. This correspondence is now available for viewing in Private PAIR.

The official date of notification of the outgoing correspondence will be indicated on the form PTOL-90 accompanying the correspondence.

Disclaimer:
The list of documents shown below is provided as a courtesy and is not part of the official file wrapper. The content of the images shown in PAIR is the official record.

Application	Document	Mailroom Date	Attorney Docket No.
14512414	APP.FILE.REC	10/20/2014	ED 2000

To view your correspondence online or update your email addresses, please visit us anytime at https://sportal.uspto.gov/secure/myportal/privatepair.

If you have any questions, please email the Electronic Business Center (EBC) at EBC@uspto.gov with 'e-Office Action' on the subject line or call 1-866-217-9197 during the following hours:

Monday - Friday 6:00 a.m. to 12:00 a.m.

Thank you for prompt attention to this notice,

UNITED STATES PATENT AND TRADEMARK OFFICE

PATENT APPLICATION INFORMATION RETRIEVAL SYSTEM

RED HAT 1010
RED HAT V. CAS

PATENT APPLICATION FEE DETERMINATION RECORD

Substitute for Form PTO-875

Application or Docket Number: 14/512,414

APPLICATION AS FILED - PART I

FOR	(Column 1) NUMBER FILED	(Column 2) NUMBER EXTRA	SMALL ENTITY RATE($)	SMALL ENTITY FEE($)	OR	OTHER THAN SMALL ENTITY RATE($)	OTHER THAN SMALL ENTITY FEE($)
BASIC FEE (37 CFR 1.16(a), (b), or (c))	N/A	N/A	N/A	70		N/A	
SEARCH FEE (37 CFR 1.16(k), (l), or (m))	N/A	N/A	N/A	300		N/A	
EXAMINATION FEE (37 CFR 1.16(o), (p), or (q))	N/A	N/A	N/A	360		N/A	
TOTAL CLAIMS (37 CFR 1.16(i))	20 minus 20=	*	x 40 =	0.00	OR		
INDEPENDENT CLAIMS (37 CFR 1.16(h))	4 minus 3 =	1	x 210 =	210			
APPLICATION SIZE FEE (37 CFR 1.16(s))	If the specification and drawings exceed 100 sheets of paper, the application size fee due is $310 ($155 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).			0.00			
MULTIPLE DEPENDENT CLAIM PRESENT (37 CFR 1.16(j))				0.00			
* If the difference in column 1 is less than zero, enter "0" in column 2.			TOTAL	940		TOTAL	

APPLICATION AS AMENDED - PART II

		(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	SMALL ENTITY RATE($)	SMALL ENTITY ADDITIONAL FEE($)	OR	OTHER THAN SMALL ENTITY RATE($)	OTHER THAN SMALL ENTITY ADDITIONAL FEE($)
AMENDMENT A	Total (37 CFR 1.16(i))	*	Minus	**	=	x =		OR	x =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=	x =		OR	x =	
	Application Size Fee (37 CFR 1.16(s))									
	FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							OR		
						TOTAL ADD'L FEE		OR	TOTAL ADD'L FEE	

		(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	SMALL ENTITY RATE($)	SMALL ENTITY ADDITIONAL FEE($)	OR	OTHER THAN SMALL ENTITY RATE($)	OTHER THAN SMALL ENTITY ADDITIONAL FEE($)
AMENDMENT B	Total (37 CFR 1.16(i))	*	Minus	**	=	x =		OR	x =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=	x =		OR	x =	
	Application Size Fee (37 CFR 1.16(s))									
	FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							OR		
						TOTAL ADD'L FEE		OR	TOTAL ADD'L FEE	

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.

** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20".

*** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3".

The "Highest Number Previously Paid For" (Total or Independent) is the highest found in the appropriate box in column 1.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING or 371(c) DATE	GRP ART UNIT	FIL FEE REC'D	ATTY.DOCKET.NO	TOT CLAIMS	IND CLAIMS
14/512,414	10/11/2014	2652	940	ED 2000	20	4

CONFIRMATION NO. 9530

64828
Harold McGurk
The Law Office of Clay McGurk
P.O. BOX 1488
Orange, CA 92856

FILING RECEIPT

OC00000007139039l

Date Mailed: 10/23/2014

Receipt is acknowledged of this non-provisional patent application. The application will be taken up for examination in due course. Applicant will be notified as to the results of the examination. Any correspondence concerning the application must include the following identification information: the U.S. APPLICATION NUMBER, FILING DATE, NAME OF APPLICANT, and TITLE OF INVENTION. Fees transmitted by check or draft are subject to collection. Please verify the accuracy of the data presented on this receipt. **If an error is noted on this Filing Receipt, please submit a written request for a Filing Receipt Correction. Please provide a copy of this Filing Receipt with the changes noted thereon. If you received a "Notice to File Missing Parts" for this application, please submit any corrections to this Filing Receipt with your reply to the Notice. When the USPTO processes the reply to the Notice, the USPTO will generate another Filing Receipt incorporating the requested corrections**

Inventor(s)

Eric M. Delangis, Huntsville, AL;

Applicant(s)

Eric M. Delangis, Huntsville, AL;

Power of Attorney: The patent practitioners associated with Customer Number 64828

Domestic Priority data as claimed by applicant

This application is a CON of 13/531,294 06/22/2012 PAT 8861349 *
which is a CON of 12/581,852 10/19/2009 PAT 8228801
which is a CON of 10/686,375 10/14/2003 PAT 7606156
which claims benefit of 60/418,521 10/15/2002
(*)Data provided by applicant is not consistent with PTO records.

Foreign Applications for which priority is claimed (You may be eligible to benefit from the **Patent Prosecution Highway** program at the USPTO. Please see http://www.uspto.gov for more information.) - None.
Foreign application information must be provided in an Application Data Sheet in order to constitute a claim to foreign priority. See 37 CFR 1.55 and 1.76.

If Required, Foreign Filing License Granted: 10/16/2014
The country code and number of your priority application, to be used for filing abroad under the Paris Convention, is **US 14/512,414**
Projected Publication Date: 01/29/2015
Non-Publication Request: No
Early Publication Request: No
**** SMALL ENTITY ****

RED HAT 1010
RED HAT V. CAS

Title

MULTIPATH COMMUNICATION DEVICES AND METHODS

Preliminary Class

379

Statement under 37 CFR 1.55 or 1.78 for AIA (First Inventor to File) Transition Applications: No

PROTECTING YOUR INVENTION OUTSIDE THE UNITED STATES

Since the rights granted by a U.S. patent extend only throughout the territory of the United States and have no effect in a foreign country, an inventor who wishes patent protection in another country must apply for a patent in a specific country or in regional patent offices. Applicants may wish to consider the filing of an international application under the Patent Cooperation Treaty (PCT). An international (PCT) application generally has the same effect as a regular national patent application in each PCT-member country. The PCT process **simplifies** the filing of patent applications on the same invention in member countries, but **does not result** in a grant of "an international patent" and does not eliminate the need of applicants to file additional documents and fees in countries where patent protection is desired.

Almost every country has its own patent law, and a person desiring a patent in a particular country must make an application for patent in that country in accordance with its particular laws. Since the laws of many countries differ in various respects from the patent law of the United States, applicants are advised to seek guidance from specific foreign countries to ensure that patent rights are not lost prematurely.

Applicants also are advised that in the case of inventions made in the United States, the Director of the USPTO must issue a license before applicants can apply for a patent in a foreign country. The filing of a U.S. patent application serves as a request for a foreign filing license. The application's filing receipt contains further information and guidance as to the status of applicant's license for foreign filing.

Applicants may wish to consult the USPTO booklet, "General Information Concerning Patents" (specifically, the section entitled "Treaties and Foreign Patents") for more information on timeframes and deadlines for filing foreign patent applications. The guide is available either by contacting the USPTO Contact Center at 800-786-9199, or it can be viewed on the USPTO website at http://www.uspto.gov/web/offices/pac/doc/general/index.html.

For information on preventing theft of your intellectual property (patents, trademarks and copyrights), you may wish to consult the U.S. Government website, http://www.stopfakes.gov. Part of a Department of Commerce initiative, this website includes self-help "toolkits" giving innovators guidance on how to protect intellectual property in specific countries such as China, Korea and Mexico. For questions regarding patent enforcement issues, applicants may call the U.S. Government hotline at 1-866-999-HALT (1-866-999-4258).

LICENSE FOR FOREIGN FILING UNDER

Title 35, United States Code, Section 184

Title 37, Code of Federal Regulations, 5.11 & 5.15

GRANTED

The applicant has been granted a license under 35 U.S.C. 184, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" followed by a date appears on this form. Such licenses are issued in all applications where the conditions for issuance of a license have been met, regardless of whether or not a license may be required as set forth in 37 CFR 5.15. The scope and limitations of this license are set forth in 37 CFR 5.15(a) unless an earlier license has been issued under 37 CFR 5.15(b). The license is subject to revocation upon written notification. The date indicated is the effective date of the license, unless an earlier license of similar scope has been granted under 37 CFR 5.13 or 5.14.

This license is to be retained by the licensee and may be used at any time on or after the effective date thereof unless it is revoked. This license is automatically transferred to any related applications(s) filed under 37 CFR 1.53(d). This license is not retroactive.

The grant of a license does not in any way lessen the responsibility of a licensee for the security of the subject matter as imposed by any Government contract or the provisions of existing laws relating to espionage and the national security or the export of technical data. Licensees should apprise themselves of current regulations especially with respect to certain countries, of other agencies, particularly the Office of Defense Trade Controls, Department of State (with respect to Arms, Munitions and Implements of War (22 CFR 121-128)); the Bureau of Industry and Security, Department of Commerce (15 CFR parts 730-774); the Office of Foreign AssetsControl, Department of Treasury (31 CFR Parts 500+) and the Department of Energy.

NOT GRANTED

No license under 35 U.S.C. 184 has been granted at this time, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" DOES NOT appear on this form. Applicant may still petition for a license under 37 CFR 5.12, if a license is desired before the expiration of 6 months from the filing date of the application. If 6 months has lapsed from the filing date of this application and the licensee has not received any indication of a secrecy order under 35 U.S.C. 181, the licensee may foreign file the application pursuant to 37 CFR 5.15(b).

SelectUSA

The United States represents the largest, most dynamic marketplace in the world and is an unparalleled location for business investment, innovation, and commercialization of new technologies. The U.S. offers tremendous resources and advantages for those who invest and manufacture goods here. Through SelectUSA, our nation works to promote and facilitate business investment. SelectUSA provides information assistance to the international investor community; serves as an ombudsman for existing and potential investors; advocates on behalf of U.S. cities, states, and regions competing for global investment; and counsels U.S. economic development organizations on investment attraction best practices. To learn more about why the United States is the best country in the world to develop technology, manufacture products, deliver services, and grow your business, visit http://www.SelectUSA.gov or call +1-202-482-6800.

To: CLAYMCGURK@GMAIL.COM,,
From: PAIR_eOfficeAction@uspto.gov
Cc: PAIR_eOfficeAction@uspto.gov
Subject: Private PAIR Correspondence Notification for Customer Number 64828

Oct 23, 2014 05:24:16 AM

Dear PAIR Customer:

Harold McGurk
The Law Office of Clay McGurk
P.O. BOX 1488
Orange, CA 92856
UNITED STATES

The following USPTO patent application(s) associated with your Customer Number, 64828 , have new outgoing correspondence. This correspondence is now available for viewing in Private PAIR.

The official date of notification of the outgoing correspondence will be indicated on the form PTOL-90 accompanying the correspondence.

Disclaimer:
The list of documents shown below is provided as a courtesy and is not part of the official file wrapper. The content of the images shown in PAIR is the official record.

Application	Document	Mailroom Date	Attorney Docket No.
14512414	APP.FILE.REC	10/23/2014	ED 2000

To view your correspondence online or update your email addresses, please visit us anytime at https://sportal.uspto.gov/secure/myportal/privatepair.

If you have any questions, please email the Electronic Business Center (EBC) at EBC@uspto.gov with 'e-Office Action' on the subject line or call 1-866-217-9197 during the following hours:

Monday - Friday 6:00 a.m. to 12:00 a.m.

Thank you for prompt attention to this notice,

UNITED STATES PATENT AND TRADEMARK OFFICE

PATENT APPLICATION INFORMATION RETRIEVAL SYSTEM

RED HAT 1010
RED HAT V. CAS

PTO/SB/08a (01-10)
Approved for use through 07/31/2012. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT
(Not for submission under 37 CFR 1.99)

Application Number	14512414
Filing Date	2014-10-11
First Named Inventor	Eric Delangis
Art Unit	2652
Examiner Name	
Attorney Docket Number	ED 2000

U.S.PATENTS Remove

Examiner Initial*	Cite No	Patent Number	Kind Code[1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1					

If you wish to add additional U.S. Patent citation information please click the Add button. Add

U.S.PATENT APPLICATION PUBLICATIONS Remove

Examiner Initial*	Cite No	Publication Number	Kind Code[1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1					

If you wish to add additional U.S. Published Application citation information please click the Add button. Add

FOREIGN PATENT DOCUMENTS Remove

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2] i	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	T[5]
	1							☐

If you wish to add additional Foreign Patent Document citation information please click the Add button Add

NON-PATENT LITERATURE DOCUMENTS Remove

Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc), date, pages(s), volume-issue number(s), publisher, city and/or country where published.	T[5]

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	14512414
	Filing Date	2014-10-11
	First Named Inventor	Eric Delangis
	Art Unit	2652
	Examiner Name	
	Attorney Docket Number	ED 2000

	1	Fisk et al., "Dynamic Right-Sizing in TCP", Proceedings of the Los Alamos Computer Science Institute Symposium", pp. 01-5460 (2001)	☐
	2	Hsieh et al., "pTCP: An End-to-End Transport Layer Protocol for Striped Connections", ICNP, pp. 24-33 IEEE Computer Society (2002)	☐

If you wish to add additional non-patent literature document citation information please click the Add button **Add**

EXAMINER SIGNATURE

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	14512414
	Filing Date	2014-10-11
	First Named Inventor	Eric Delangis
	Art Unit	2652
	Examiner Name	
	Attorney Docket Number	ED 2000

CERTIFICATION STATEMENT

Please see 37 CFR 1.97 and 1.98 to make the appropriate selection(s):

☐ That each item of information contained in the information disclosure statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(1).

OR

☒ That no item of information contained in the information disclosure statement was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the person signing the certification after making reasonable inquiry, no item of information contained in the information disclosure statement was known to any individual designated in 37 CFR 1.56(c) more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(2).

☐ See attached certification statement.

☐ The fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☐ A certification statement is not submitted herewith.

SIGNATURE

A signature of the applicant or representative is required in accordance with CFR 1.33, 10.18. Please see CFR 1.4(d) for the form of the signature.

Signature	/H C McGurk IV/	Date (YYYY-MM-DD)	2014-11-05
Name/Print	H C McGurk	Registration Number	34964

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 1 hour to complete, including gathering, preparing and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these record s.

2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.

3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.

4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).

5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.

6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).

7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.

8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.

9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION NO.:	**14/512,414**	EXAMINER:	
INVENTOR:	Eric M. Delangis	GROUP ART UNIT:	2652
FILED:	OCTOBER 11, 2014	CONFIRMATION NO.:	9530
TITLE:	Multipath Communication Devices and Methods		
ATTORNEY NO:	ED 2000		

Commissioner For Patents
PO Box 1450
Alexandria, VA 22313-1450

INFORMATION DISCLOSURE STATEMENT TRANSMITTAL LETTER

This submission does not represent that a search has been made or that no better art exists and does not constitute an admission that each or all of the listed documents are material or constitute prior art. If the Examiner applies any of the documents as prior art against any of the claims in the present application, all rights are reserved to take appropriate action to establish the patentability of the disclosed invention over the present application.

Best Regards,

Date: November 5, 2014

H C McGurk IV, Reg. No. 34,964
Attorney for Applicant

The Law Office of Clay McGurk
PO Box 1488
Orange, CA 92856
Phone: **949.395.3942**

CERTIFICATE OF ELECTRONIC (EFS-WEB) TRANSMISSION

I hereby certify that this correspondence is being transmitted via the U.S. Patent Office's electronic filing system in accordance with 37 C.F.R. § 1.8(a)(1) from the **Pacific Time Zone** of the United States on the local date shown below.

Dated: November 5, 2014

By: Clay McGurk

RED HAT 1010
RED HAT V. CAS

Published in the Proc. of the 14th Annual ACM/IEEE SC2001 Conference, November 2001. LA-UR 01-2823

Dynamic Right-Sizing: TCP Flow-Control Adaptation

Mike Fisk[§†] *mfisk@lanl.gov*
Wu-chun Feng[§‡] *feng@lanl.gov*

[§]Los Alamos National Laboratory*
[†]Department of Computer Science & Engineering, University of California, San Diego
[‡]Department of Computer & Information Science, The Ohio State University

1 Introduction

Grid and networking researchers continue the practice of manually optimizing TCP buffer sizes to keep the network pipe full [10, 7], and thus achieve acceptable performance over the wide-area network, whether for bulk-data transfer or in support of computational grids [4], data grids [1, 8, 3], or access grids [2]. Not only is this process cumbersome, but the result of tuning window sizes for a particular pair of hosts is sub-par performance for connections with larger delay-bandwidth products and the misappropriation of scarse resources to connections with smaller delay-bandwidth products [9].

Consequently, we propose an operating system technique called Dynamic Right-Sizing that eliminates the need for this manual process. Compared to previous work on this problem, our solution is more efficient and both more transparent and widely-usable to applications.

The Web100 project has released a modified FTP client that uses user-space code to send a burst of pings to estimate the latency and bandwidth at the beginning of a connection and adjust the windows accordingly [6]. This approach requires that this measurement code be deployed in each application. Further, it uses a measurement period before data is sent and creates extra network traffic that is not controlled by a congestion avoidance mechanism. Thus, it can only be used sparingly, such as at the beginning of a connection. In contrast, our implementation automatically benefits not just FTP, but every application on a host.

Earlier work in [9] presents kernel modifications for 'auto-tuning' a sender's flow-control window based on the congestion window and then using fair-share algorithms to manage competition between connections for buffers. In this scheme, the receiver's window advertisements are superfluous. This solution does not adequaquately solve the problem for applications such as archival storage where the bottleneck in a connection is the receiving application instead of the network. In this context, the receiver still needs flow control or else unbounded amounts of receiver buffer space can be consumed by data that has been received and acknowledged by the operating system, but is waiting for the receiving application. This could in turn require the receiver to resort to dropping packets and unnecessary triggering of the sender's congestion avoidance mechanisms, thus wasting bandwidth and reducing throughput. Our technique complements the 'auto-tuning' presented in [9] by providing the receiver with the ability to measure the size of the sender's congestion window, and more fairly allocate buffers to connections based on their need for buffers.

*This work was supported by the U.S. Dept. of Energy's Next Generation Internet - Earth Systems Grid and Accelerated Strategic Computing Initiative - Distance and Distributed Computing and Communication programs through Los Alamos National Laboratory contract W-7405-ENG-36.

2 Dynamic Right-sizing

In short, *Dynamic Right-sizing* lets the receiver estimate the sender's congestion window size and use that estimate to dynamically change the size of the receiver's window advertisements. As a result, the sender will be congestion-window-limited rather than flow-control-window-limited.

A sender can send no more than one window's worth of data between acknowledgements. Accordingly, a burst that is shorter than a round-trip time can contain at most one window's worth of data. Thus, for any period of time that is shorter than a round-trip time, the amount of data seen by the receiver over that period is a lower-bound on the size of the sender's window. Some data may be lost or delayed by the network, so the sender may have sent more than the amount of data seen. Further, the sender may not have had a full window's worth of data to send. So the window may be significantly larger than this lower-bound, but not if the connection is truly limited by the receiver's window. Measuring this minimum and making sure that the receiver's advertised window is always larger will let the receiver track the congestion window size.

To make these measurements, it is necessary for the receiver to know the round-trip time. In a typical TCP implementation, the round-trip time is measured by observing the time between when data is sent and an acknowledgement is returned [5]. But during a bulk-data transfer, the receiver might not be sending any data and would therefore not have a good round-trip time estimate. For instance, an FTP data connection transmits data entirely in one direction. A system that is only transmitting acknowledgements can still estimate the round-trip time by observing the time between when a byte is first acknowledged and the receipt of data that is at least one window beyond the sequence number that was acknowledged. If the sender is being throttled by the network, this estimate will be valid. However, if the sending application did not have any data to send, the measured time could be much larger than the actual round-trip time. Thus this measurement acts only as an upper-bound on the round-trip time and should be be used only when it is the only source of round-trip time information.

RED HAT 1010
RED HAT V. CAS

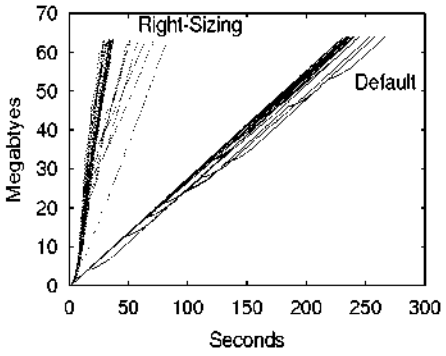


Figure 1: Progress of data transfers

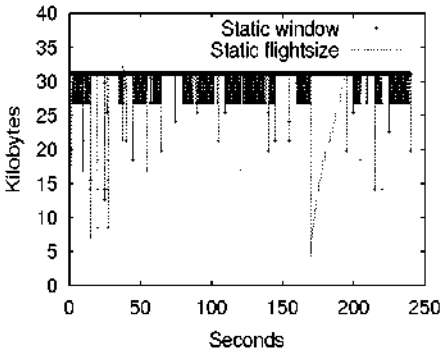


Figure 2: Default window size
Flight & window sizes

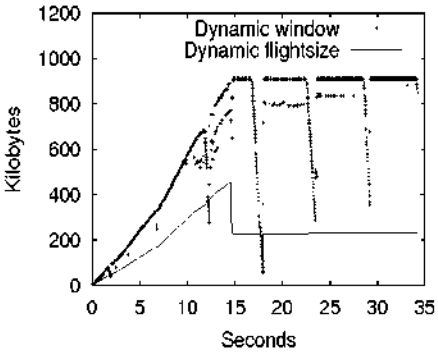


Figure 3: Dynamic right-sizing
Flight & window sizes

2.1 Impact of Timer Imprecision

Like many systems, the system timer in Linux is limited to a precision of 10ms. In this section we explore the impact of using this efficient, but imprecise measure of time.

Assume the salient scenario where a sender is always sending as fast as possible, but that the network may delay packets arbitrarily. However, we are concerned with the case were the sender is limited by the window-size, so that network delays are small enough that no timeouts occur. Further assume that the window-size is not fluctuating during a measurement. Then the receiver can determine that the window size limiting the sender is bounded as follows:

$$\frac{d}{n_{max}} \leq w \leq \frac{d}{n_{min}} \tag{1}$$

Where d bytes of data have been received over some number of round-trip times between n_{min} and n_{max}.

Any measurement period consists of some number of whole round-trip times plus fractional round-trip times preceding and following the complete round-trips. The possible number of round-trip periods observed is a whole number bounded as follows:

$$\left\lceil \frac{t}{rtt} \right\rceil \leq n < \left\lfloor \frac{t}{rtt} \right\rfloor + 2 \tag{2}$$

Due to similar fence-post problems, a measurement of duration equal to one round-trip time may actually be up to 20ms longer than the round-trip time. Combining these facts with bounds 1 and 2 yields:

$$\frac{d}{\lfloor \frac{20\text{ms}}{rtt} \rfloor + 3} \leq w \leq d \tag{3}$$

Thus, in no case will the actual window be larger than the measured amount of data received during the period. However, the amount of data received during the period may be three times the actual window size when measurements are made across wide-area networks with $rtt > 20$ms. Further, local networks with small round-trip delays may be grossly overestimated.

We therefore conclude that measurements made with coarse timers will not cause dynamic right-sizing to underestimate the window size or negatively impact throughput. However, to make more accurate decisions for memory sharing under pressure, it is advantageous to use precise timers. Many CPU architectures now feature hardware time counters that can be used to efficiently obtain a precise timestamp. In future versions of our implementation, we will pursue the use of these counters. However, our use of these counters will also require that the standard TCP round-trip time estimation be done with equivalent precision.

3 Experimental Performance

In Figure 1, 50 transfers were made between Linux 2.2 systems modified to perform Dynamic Right-Sizing and connected by a network with an emulated round-trip delay of 100ms. The first 25 transfers used the default window sizes of 64 kilobytes for both the sender and receiver. The second 25 transfers, shown in dotted lines, used the dynamically sized windows described above.

In Figures 2 and 3, we examine the window sizes during two of the above transfers. The amount of sent, but unacknowledged data in the sender's buffer is known as the flightsize. The flightsize is in turn bounded by the window advertised by the receiver. Figure 2 shows that in the traditional, static case without dynamic right-sizing, the congestion window, and consequently flightsize quickly grow equal to the size of the window advertisements. For most of the duration of the connection, it is limited by the receiver's low window advertisement of 32KB.

In contrast, during the dynamic right-sizing case shown in Figure 3, the receiver is able, during most of the connection, to advertise a window size that is roughly twice the largest flightsize seen to date. As a result, the flightsize is only constrained by the congestion window and the delay-bandwidth product. Slow start continues for much longer and stops only when there is packet loss. At this point the congestion window stabilizes on a flightsize that is 7 times higher than the constrained flightsize of the static case. This 7-fold increase in the average flightsize is the source of the same, 7-fold increase in throughput demonstrated in Figure 1.

References

[1] Ann Chervenak, Ian Foster, Carl Kesselman, Charles Salisbury, and Steven Tuecke, "The data grid: Towards an

RED HAT 1010
RED HAT V. CAS

architecture for the distributed management and analysis of large scientific datasets," *Journal of Network and Computer Applications*, vol. 23, no. 3, pp. 187–200, July 2000.

[2] Lisa Childers, Terry Disz, Robert Olson, Michael E. Papka, Rick Stevens, and Tushar Udeshi, "Access grid: Immersive group-to-group collaborative visualization," in *Proceedings of the 4th International Immersive Projection Technology Workshop*, 2000.

[3] "EU-data grid," http://www.eu-datagrid.org/.

[4] Ian Foster and Carl Kesselman, *The Grid: Blueprint for a New Computing Infrastructure*, Morgan-Kaufmann, 1998.

[5] Van Jacobson, "Congestion Avoidance and Control," in *Proceedings, SIGCOMM '88 Workshop*. ACM SIGCOMM, Aug. 1988, pp. 314–329, ACM Press, Stanford, CA.

[6] Jian Lui and Jim Ferguson, "Automatic TCP socket buffer tuning," in *Supercomputing 2000 Research Gems*, Nov. 2000, Awarded "Best Research Gem of the Conference," http://dast.nlanr.net/Features/Autobuf/.

[7] Jamshid Mahdavi, "Enabling high performance data transfers on hosts," Webpage, http://www.psc.edu/networking/perf_tune.html.

[8] "Particle physics data grid," http://www.cacr.caltech.edu/ppdg/.

[9] Jeff Semke, Jamshid Mahdavi, and Matt Mathis., "Automatic TCP buffer tuning," *Computer Communications Review*, vol. 28, no. 4, pp. 315–323, Oct. 1998.

[10] Brian L. Tierney, "TCP tuning guide for distributed application on wide area networks," *:login*, vol. 26, no. 1, Feb. 2001.

RED HAT 1010
RED HAT V. CAS

pTCP: An End-to-End Transport Layer Protocol for Striped Connections *

Hung-Yun Hsieh and Raghupathy Sivakumar
School of Electrical and Computer Engineering
Georgia Institute of Technology
Email: {hyhsieh, siva}@ece.gatech.edu

Abstract

The TCP transport layer protocol is designed for connections that traverse a single path between the sender and receiver. However, there are several environments in which multiple paths can be used by a connection simultaneously. In this paper we consider the problem of supporting striped connections that operate over multiple paths. We propose an end-to-end transport layer protocol called pTCP that allows connections to enjoy the aggregate bandwidths offered by the multiple paths, irrespective of the individual characteristics of the paths. We show that pTCP can have a varied range of applications through instantiations in three different environments: (a) bandwidth aggregation on multi-homed mobile hosts, (b) service differentiation using purely end-to-end mechanisms, and (c) end-systems based network striping. In each of the applications we demonstrate the applicability of pTCP and how its efficacy compares with existing approaches through simulation results.

1. Introduction

The TCP transport layer protocol is designed for connections that traverse a single path between the source and the destination. If a TCP connection is striped over multiple network paths, the performance of the aggregate connection can potentially be even worse than the bandwidth available along any of the paths. This is because TCP expects a predominantly first-in-first-out delivery of packets through the network, and uses out-of-order arrival of packets as indication of packet losses due to congestion. A seemingly better solution of using *application layer striping* over multiple TCP sockets [2, 6, 12] (one each for every path available), does not fare well either due to a variety of reasons, the primary one being the inability to accurately profile bandwidths available along individual paths at the application layer. While we discuss the performance of the application layer striping approach later in the paper, we contend that TCP as-is is not an appropriate transport protocol for striped connections.

On the other hand, striping of connections can have considerable benefits in a variety of Internet settings. For example, a multi-homed mobile host can stripe connections on its multiple wireless interfaces, and thus achieve aggregate bandwidths in a typically bandwidth scarce environment. Striping of connections on the backbone Internet has potential advantages including better utilization of network resources and increased fairness to the connections. While there are several such gains to be attained through striping, the focus of this paper is not to motivate connection striping. Instead, we consider the question: *If connections are to be striped, how can a transport protocol support such connections effectively?* We measure effectiveness as the ability of the transport protocol to deliver the aggregate bandwidths available along the multiple paths to the application. A subset of the challenges such a transport protocol has to address include: (a) the different paths available to the striped connection can have diverse characteristics in terms of bandwidth, delay, and loss; (b) path characteristics can be prone to both short-term and long-term fluctuations; (c) in the worst case, a path can experience a complete shutdown; and (d) the buffer at the receiver is finite irrespective of whether it is partitioned evenly among the different paths or is shared more dynamically.

In this context, we propose an end-to-end transport layer protocol called *pTCP* (parallel TCP) that effectively supports striped connections. Although the pTCP design does not require any specific behavior from the component flows of the striped connection, in this paper we focus only on the case where a component flow exhibits the same behavior as that of a regular TCP connection. pTCP achieves effective aggregation of bandwidth for a striped connection through a combination of unique mechanisms including: (a) decoupling functionalities pertaining to the aggregate connection from those that pertain to an individual path; (b) effective striping across the multiple paths based on the in-

*This work was supported in part by funds from NSF (award #ANI-0117840) and Yamacraw (http://www.yamacraw.org).

RED HAT 1010
RED HAT V. CAS

stantaneous bandwidths; and (c) appropriate re-striping and redundant striping of packets during periods of fluctuation in path characteristics.

As pointed out earlier, a protocol such as pTCP can have applications in several different settings. We identify and focus on the following three application domains for pTCP: (a) bandwidth aggregation at mobile hosts with multiple wireless interfaces, (b) end-to-end weighted rate service differentiation, and (c) striping on an overlay network based on application layer switching. For each of the settings, we discuss the applicability of pTCP, why pTCP is superior to existing approaches in each of the domains, and present simulation results that substantiate our arguments.

The contributions of this work are thus twofold:

- *We propose an end-to-end transport protocol called pTCP for striped connections. pTCP uses a combination of mechanisms to provide a striped connection with the aggregate bandwidth available along the multiple paths.*
- *We also show the use of pTCP in three different Internet settings, and demonstrate its efficacy when compared to existing approaches in those settings.*

The rest of the paper is organized as follows: Section 2 lists the key design principles that underlie pTCP. Section 3 presents the software architecture and protocol overview of pTCP. Section 4 discusses the three application domains for pTCP. Finally, Section 5 discusses some related work and concludes the paper. In the rest of the section, we define some of the key terminology that we use in the paper.

1.1. Terminology

- *Striped Connection:* We refer to a connection that has multiple delivery paths between the source and the destination as a striped connection.
- *Micro-Flow:* We refer to the portion of a striped connection that traverses a single individual path as a component flow or a micro-flow. Hence, several micro-flows together comprise the striped connection. In this paper, we assume that a micro-flow will exhibit the same behavior as a normal TCP connection.
- *Aggregate Bandwidth:* We refer to the sum of the bandwidths available for each of the micro-flows as the aggregate bandwidth. Note that the challenge of providing the aggregate bandwidth to the striped connection is non-trivial.

2. pTCP Design

The key obstacle to achieving the aggregate bandwidth for striped connections is that each of the individual paths can have vastly differing characteristics in terms of bandwidth and delay (round-trip time). If data-striping is done without taking into account these differences, the bandwidth achieved by the striped connection can be significantly lower than the maximum possible. For example, consider a connection striped across two paths $p1$ and $p2$ with bandwidths of $b1$ and $b2$ respectively (where $b1 < b2$), and the same delay d. If data is striped across the two paths equally, the bandwidth achieved is theoretically $(2 * b1)$ which is smaller than the ideal bandwidth of $(b1 + b2)$.

However, in TCP the achieved bandwidth can be even smaller. Consider a fast micro-flow $f2$ co-existing with a slower micro-flow $f1$. If data is striped equally, $f2$ is likely to experience an overflow of its receive buffer (since packets will be buffered waiting for in-sequence packets to arrive through $f1$). In TCP, when a window advertisement of zero is made, the sender enters the *persist state* and can take a considerable amount of time to recover (either through an explicit window update sent by the receiver, or through the expiry of TCP's persist timer that has a starting value of 5 seconds, and a maximum of 60 seconds [17]). In fact, this can induce a domino effect of micro-flows alternately entering persist states if any of the explicit updates are lost. Not surprisingly, a simple scheme that stripes data down a path till the send buffer of the micro-flow blocks does not perform well either, as long as the size of the send buffer is different from the bandwidth-delay product of the micro-flow (which is typically true in TCP).

The problems that arise due to bandwidth differences can be solved by making sure that the data-striping ratio is the same as the ratio of the bandwidths of the different paths. From the standpoint of TCP, this translates into striping data based on the ratio of the $\frac{congestion\_window\_size}{round\_trip\_time}$ values for the different micro-flows. The problems that arise due to delay differences, on the other hand, can be solved by simply over-allocating the receive buffer of the striped connection. Specifically, for k paths where path p_i has bandwidth b_i and delay d_i, a total buffer allocation of $(\Sigma b_i) * max(d_i)$ (as opposed to $\Sigma(b_i * d_i)$) will absorb the problems caused by delay differences. Similar observations have also been made in [9]. In the rest of the paper, we assume the use of such buffer provisioning at the receiver to overcome delay differences.

The problems identified this far are further exacerbated when fluctuations of path characteristics occur. For example, when the bandwidth of a path p_i decreases, for a transient period the number of packets outstanding on the path is disproportional to its bandwidth. If bandwidth fluctuations are a norm (like in wireless environments), this can result in a failure to keep the striping ratio the same as the ratio of the bandwidths. In the rest of the section, we list the fundamental design elements in pTCP that are targeted toward addressing the problems discussed above.

RED HAT 1010
RED HAT V. CAS

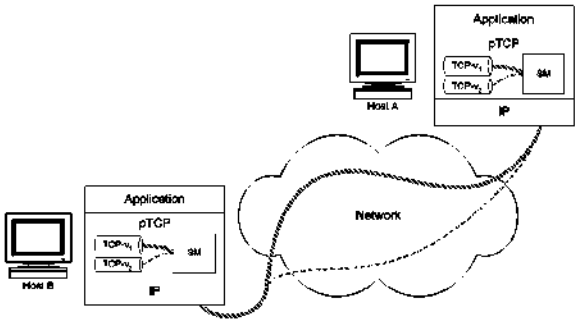


Figure 1. Use pTCP for Striped Connections

- *Decoupling of Functionalities:* pTCP decouples functionalities associated with per-path characteristics from those that pertain to the aggregate connection. We refer to the component in pTCP that handles aggregate connection functionalities as *SM* (Striped connection Manager), and the component in pTCP that handles per-path functionalities as *TCP-v* (TCP-virtual). TCP-v is a modified version of TCP that deals only with virtual packets and virtual buffers.[1] Each micro-flow of a striped connection is controlled by an independent TCP-v.

  Congestion control, the mechanism that estimates the available bandwidth along a path, is obviously a per-path functionality and is handled by TCP-v. Since SM is responsible for striping data across the different TCP-vs, it handles *buffer management* (including sequencing at the receiver), and consequently *flow control*. *Reliability* in pTCP is primarily handled by TCP-v just as in TCP. However, under certain circumstances that we describe later, SM can intervene and cause loss recovery to occur along a path different from the one on which the loss occurred. Finally, *connection management* is handled by SM, although it recursively uses the connection management performed by individual TCP-vs. Figure 1 illustrates the roles of the two components in pTCP.

- *Delayed Binding:* Although striping based on the $\frac{cwnd}{rtt}$ ratios is mostly sufficient to achieve aggregate bandwidths, such a scheme will remain vulnerable to instantaneous fluctuations in the bandwidth and delay of a path. pTCP uses a delayed binding strategy that satisfies the requirement of striping based on ratios of $\frac{cwnd}{rtt}$, and at the same time dynamically adapts to instantaneous fluctuations in bandwidth and delay. Specifically, since the buffer management is performed by SM, TCP-v operates only on virtual buffers. Thus, when a packet is transmitted by TCP-v, the virtual packet should be bound to the real data in the buffer at SM. pTCP performs this binding only just before the virtual packet is transmitted. This coupled with TCP's self-clocked property results in pTCP adapting the data-striping to instantaneous changes in bandwidth and delay. Looking at it from another perspective, any packet bound to any TCP-v will have already been transmitted or will be in the process of being transmitted, precluding any chances of packets being held up in buffers at the source due to bandwidth fluctuations. An exception to this rule is when there is a packet loss and the lost packet falls outside the reduced congestion window of the concerned TCP-v. In this case, the lost packet is bound to a particular TCP-v, but is not in transit through the network. We describe how pTCP handles such a situation in the next design element.

- *Packet Re-striping:* In steady state, pTCP will ensure that the number of outstanding packets in a micro-flow is proportional to the bandwidth along the corresponding path. Moreover, the delayed binding strategy further ensures that all bound packets are already in transit in the network. However, during congestion when packets are lost in the network, the reduction of the congestion window by a TCP-v can result in bound packets falling outside the congestion window. If such packets are lost during the congestion, then they will remain un-transmitted till the congestion window of that TCP-v expands beyond their sequence numbers. This can potentially result in an overflow of the receive buffer if the other micro-flows are active in the meantime, finally resulting in a connection stall. pTCP handles the problem by unbinding packets that fall outside of the congestion window of TCP-v they were assigned to. Such unbinding results in those packets being available to be re-assigned to the next TCP-v that can send more data.

  We thus define re-striping as the process of sending a packet that was earlier transmitted through a micro-flow i, through another micro-flow j, where $j \neq i$. Note that such re-striping can presumably re-transmit data that has already been successfully delivered by the old TCP-v. This can very well be construed as an undesirable overhead. However, recall that such re-striping will be performed only during bandwidth fluctuations. If such fluctuations are rare (say the only losses are due to probe losses), then the re-striping overhead is insignificant. On the other hand, if such fluctuations are frequent, the benefits of the re-striping far outweigh the overheads incurred due to any unnecessary re-transmissions. *The results that we present in Sec-*

[1] We refer to a packet with the data portion stripped off as a virtual packet, and a buffer of virtual packets as a virtual buffer.

RED HAT 1010
RED HAT V. CAS

tion 4 show considerable performance improvement for pTCP even after accounting for the re-striping overheads.

- *Redundant Striping:* The re-striping mechanism described earlier is equivalent to a *move* operation that shifts a packet from one TCP-v to another. However, note that a packet that still remains inside the congestion window of a TCP-v will not by re-striped by SM. Consider a scenario in which a path has shutdown completely (e.g. blackout of a wireless channel). In this case, the first packet in the congestion window of the concerned TCP-v will never be re-assigned, thus potentially resulting in a connection stall. pTCP handles such cases by *redundantly striping the only packet inside the congestion window through another micro-flow* every time a timeout is experienced. A *copy* operation is performed (the packet is bound to the new micro-flow) as opposed to a *move* operation because the old TCP-v needs at least one packet to probe for recovery from the shutdown.

- *Selective Acknowledgements:* Although pTCP does not stipulate the use of any specific congestion control or reliability mechanism at TCP-v, the use of SACK information by TCP-v can significantly improve the striped connection's performance. While re-striping helps in recovering faster from loss of packets that have fallen outside the reduced congestion window of a TCP-v, SM does not intervene in the case of a lost packet that lies inside the congestion window. If there are multiple losses within the congestion window, it can take TCP-v multiple round-trip times to recover from the losses (TCP-Reno or TCP-NewReno will recover from one loss every round-trip time). This can result in the receive buffer overflow problem identified earlier. Using SACK information will enable recovering from the losses in a much shorter time span. Moreover, the availability of SACK information can also be used for more intelligent re-striping decisions and reducing the associated overheads. We do not delve into a SACK based optimization due to lack of space.

3. pTCP Architecture

3.1. Software Architecture

Figure 2 illustrates an overview of the pTCP architecture. Conceptually, SM functions as a wrapper around TCP-v. SM maintains the socket buffers and handles communication with both the higher and the lower layers. Any packet from the higher layer (say application) is queued onto the `send buffer` awaiting further treatment. Similarly, any packet from the lower layer (say IP) is queued onto the

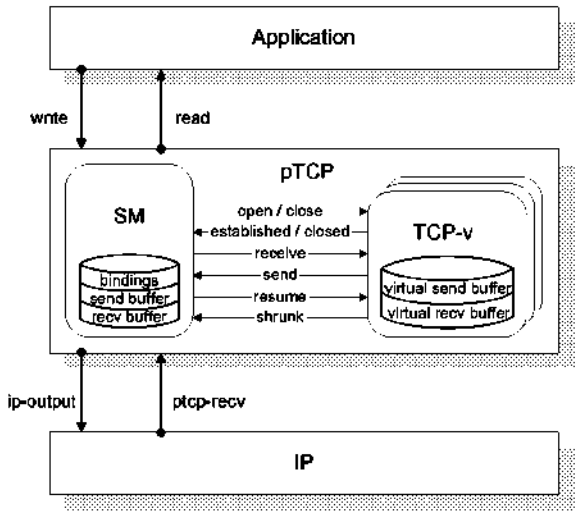


Figure 2. pTCP Architecture Overview

`recv buffer` based on its sequence number. Since the packets are handled wholly within SM, TCP-v merely operates on virtual packets and hence virtual buffers.

The interactions between SM and TCP-v that are initiated by the former are: (a) *open*, (b) *close*, (c) *receive*, and (d) *resume*. The interactions initiated by TCP-v are: (a) *established*, (b) *closed*, (c) *send*, and (d) *shrunk*. When SM receives a packet, it processes the packet, strips the data, enqueues the data onto the receive buffer, and sends the skeletal packet (sans the data) to the appropriate TCP-v through the *receive* interface. When TCP-v receives the packet, it enqueues the packet onto its virtual buffer, updates its local state, and sends any required acknowledgement packets. If a TCP-v has space in its congestion window to send new data, it invokes the *send* function with the appropriate virtual packet. SM on receiving the call binds the virtual packet to the next in-sequence unsent data packet, maintains the binding in `bindings`, and sends it out. If SM does not have new data to send, it responds with a `FREEZE` message to the concerned TCP-v. TCP-v continues to invoke the *send* function till either there is no more space left in its congestion window, or it receives a `FREEZE` from SM. In either case, the concerned TCP-v goes to sleep. SM maintains a list of *active* micro-flows consisting of those to which it has issued a `FREEZE`. When new data is received from the application, SM uses the *resume* message to notify all active micro-flows. The micro-flows then initiate the send requests as before. If a TCP-v had gone to sleep due to running out of free congestion window space, it would be awoken through a *receive* call that can potentially open up more space.

The *open* and *close* functions are used by SM to recursively invoke the *open* and *close* functions of TCP-v dur-

RED HAT 1010
RED HAT V. CAS

ing connection setup and teardown. The *established* and *closed* calls are used by TCP-v to inform SM of successfully accomplished *open* and *close* requests. The *shrunk* call is used by TCP-v to inform SM about any reduction in the congestion window size and the consequent dropout of packets from within the window.

3.2. Protocol Overview

- *Connection Management:* We assume that the number of TCP-v micro-flows to open is conveyed to pTCP when the application opens the pTCP socket. The connection setup involves recursively invoking the setup of the TCP-v micro-flows. pTCP allows data transfer to begin as soon as one of the micro-flows reaches the established state in the TCP state diagram. Similarly, for connection teardown, the close command is issued to all TCP-vs. Unlike for open, pTCP waits for all the micro-flows to be closed before returning the closed status to the application.

- *Congestion Control:* Congestion control is the sole responsibility of TCP-v. Although pTCP does not require a specific congestion control protocol to be used for a micro-flow, the discussions in this paper assume the use of TCP's congestion control scheme. One of the salient features of pTCP is the decoupling of the SM functionality from that of TCP-v. This in turn allows independent and tailored congestion control schemes to be used along the different paths.

- *Reliability:* pTCP primarily relies on the individual TCP-vs to support reliability. A packet once bound to a TCP-v is expected to be delivered reliably. The one exception to this rule is when SM unbinds a packet and re-stripes it on a different path. In this scenario, when (and if) the old micro-flow attempts a re-transmission of the packet, a new packet might be bound to the transmission. Therefore, a re-transmission at TCP-v does not need to be a re-transmission for the striped connection. This is possible because of the decoupling of functionalities between SM and TCP-v.

- *Flow Control:* Since TCP-v operates only with virtual buffers, it does not need to perform flow control. Hence, our current implementation of pTCP uses the window advertisement field already available in the TCP header for performing flow control at SM.[2] Because the overall available buffer size is an upper bound on the number of transmissions a single TCP-v can perform, the TCP flow control mechanisms do not interfere with the TCP-v operations even if such overloading of the window field is done. Note that all TCP-vs see the same advertisement and might individually attempt to transmit (subject to congestion window availability) enough data to fill in the advertised buffer space. SM, being a single point of departure for all transmissions, tackles this problem by sending a FREEZE message to the concerned TCP-vs once it knows that the total number of outstanding packets is equal to the advertised window space.

[2]Note that window scaling may be necessary since the buffer size is multiple times that of a single regular TCP connection's buffer size.

Due to lack of space we do not present the pTCP state diagram, handshakes during connection setup and teardown, and packet header formats, but instead refer interested readers to [7] for more information on these subjects.

4. pTCP Instantiations

In this section, we investigate the applicability of pTCP in the following three settings: (a) bandwidth aggregation at a mobile host with multiple wireless interfaces, (b) end-to-end approach to weighted rate differentiation, and (c) striping on an overlay network constructed through application layer switching. For each of the application domains, we discuss the key characteristics of the problem, the applicability of pTCP, and simulation results that compare pTCP to existing solutions.

4.1. Multi-homed Striping

A mobile user today has a variety of options in terms of the wireless access technologies that can be used for Internet connectivity. It is very likely that a mobile device has multiple interfaces connecting to two or more of such options. For example, a scenario in which the mobile device has a wireless LAN interface along with a wireless WAN interface is very conceivable. When such multi-homed hosts exist, an interesting question is: how are the different interfaces used in tandem?

Vertical handoffs schemes have been proposed wherein the mobile device is handed off from one access network to another depending upon its location [14]. For instance, the wireless LAN interface can be used as long as the device is inside the building. When the device moves outside the building, a vertical handoff to the wireless WAN network can be performed. Note that in the above scenario, it is possible for the mobile device to be connected to *both interfaces* when the device remains inside the building. However, an important issue is whether a connection that is striped across the two interfaces can indeed provide the aggregate bandwidth. If effective bandwidth aggregation can be achieved, connections to or from the mobile device can enjoy significantly better bandwidth performance,

RED HAT 1010
RED HAT V. CAS

especially when the bandwidths across the different interfaces are of the same order.

Link layer striping approaches such as [13] do not perform well in this context because the multiple interfaces usually belong to different access networks, and an ideal striping algorithm will need to address fluctuations in capacity caused by the end-to-end multi-hop nature of the paths. On the other hand, bandwidth-based transport layer striping approaches [9] that rely on explicit bandwidth probing of individual paths to obtain accurate estimation and thus to stripe accordingly may also suffer when aggressive probing cannot be performed (e.g. due to bandwidth constraints) or when fluctuations of link bandwidths occur at a finer time-scale than the probing period.

4.1.1. Applicability

Several properties of pTCP make it applicable to the above problem setting. pTCP uses a delayed binding strategy that handles short-term bandwidth variations common in wireless environments. The re-striping mechanism used by pTCP overcomes frequent but more permanent bandwidth variations due to reasons such as congestion. The decoupling of congestion control performed by the TCP-v from the other mechanisms at SM through a well-defined interface enables use of tailored congestion control protocols for each wireless link [7]. Finally, pTCP explicitly addresses potentially blackouts for micro-flows through its redundant striping technique.

4.1.2. Performance

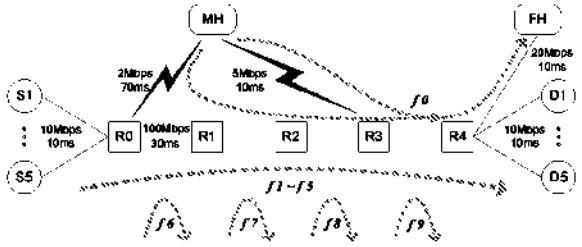


Figure 3. Network Topology for Multi-homed Striping

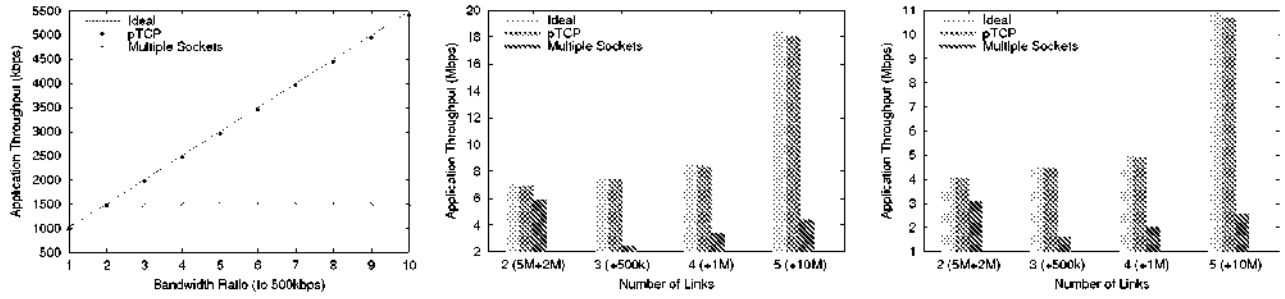
In this section, we present the simulation results showing the performance of pTCP when used in a multi-homed mobile host for bandwidth aggregation. We use the *ns-2* network simulator [15] for the network topology shown in Figure 3. For simplicity the backbone routers also double up as wireless access points (or base stations) that the mobile host (MH) can use to communicate with a remote fixed host (FH). The mobile host primarily uses two different types of wireless connections: (a) link MH-R0 with bandwidth ranging from 500Kbps to 2Mbps (depending upon the scenario) and a 70ms access delay, and (b) link MH-R3 with bandwidth ranging from 500Kbps to 5Mbps (depending upon the scenario) and a 10ms access delay. Together with the propagation delay experienced in the backbone network, these values simulate a WWAN connection with 400ms round-trip time and a WLAN connection with 100ms round-trip time respectively. MH utilizes both links using either application layer striping over multiple TCP sockets (data is striped through each socket till the socket buffer fills up) or pTCP. Besides the striped connection between MH and FH, we use 9 other flows to simulate the background traffic: 5 long TCP flows with a round-trip time of 280ms, and 4 short UDP flows with 10Mbps data rate and 100ms round-trip time as shown in the figure. We use TCP-SACK for all TCP flows including pTCP micro-flows. The simulation is run for 600 seconds and the results are obtained by averaging 10 runs. We measure the throughput delivered to the application at the destination (FH) as the metric to evaluate pTCP's performance. We also present an "ideal" aggregate throughput that is obtained by summing the throughputs of two independent TCP flows (used by two different applications) for the same scenario.

Figure 4(a) shows the simulation results when the bandwidth of the WWAN link is 500Kbps and the bandwidth of the WLAN link varies from 500Kbps to 5Mbps. pTCP is able to achieve the aggregate throughput even when the bandwidth ratio between two links goes to 10. However, by using multiple sockets the MH cannot enjoy the aggregate throughput when the ratio is greater than 2. *This result reinforces our argument that using multiple sockets cannot achieve effective bandwidth aggregation as discussed in Section 2.*

Figure 4(b) compares the performance of pTCP and multiple sockets when the number of wireless links increases from 2 to 5. Besides the 2Mbps link with 400ms round-trip time (MH-R0)and 5Mbps link with 100ms round-trip time (MH-R3), in sequence we add a 500Kbps link with 300ms round-trip time (MH-R1), a 1Mbps link with 200ms round-trip time (MH-R2) and a 10Mbps link with 50ms round-trip time (MH-R4). The bandwidth of the access link to FH is increased to 20Mbps to absorb the traffic from MH. As shown in the figure, pTCP is able to achieve effective bandwidth aggregation for multiple links. On the other hand, the aggregated bandwidth by using multiple sockets is throttled by the slowest 500Kbps link (the aggregate throughput is approximately equal to $500\text{Kbps} * (2 * 4 + 1) = 4500\text{Kbps}$ using 5 links).

Figure 4(c) shows the simulation results when all wireless links used in Figure 4(b) experience bandwidth fluctuations (say, due to channel dynamics). We simulate the scenario when the bandwidth of individual link randomly

RED HAT 1010
RED HAT V. CAS



(a) Rate Differential (b) Multiple Links (Stable) (c) Multiple Links (Fluctuating)

Figure 4. pTCP Performance for Multi-homed Striping

fluctuates between 20% to 100% of the normal values. The fluctuating period of each wireless link is set to 10, 3, 6, 20, and 5 seconds respectively. Under such fluctuating conditions, the performance of pTCP still closely follows that of the "ideal" bandwidth aggregation.

4.2. Weighted Rate Differentiation

Relative end-to-end service (rate) differentiation involves providing connections with rates relative to pre-assigned weights. For example, if two connections (with potentially different source-destination pairs) with weights $w1$ and $w2$ share the same bottleneck, their respective achieved rates $r1$ and $r2$ should be in the ratio $w1$:$w2$. Although several QoS architectures and service provisioning schemes have been proposed to address this problem, the challenge lies in achieving the rate differentiation using *purely end-to-end techniques* without relying on underlying network infrastructure.

There are two key lines of approaches for achieving end-to-end service differentiation. First, a connection with weight w is actually composed of w micro-flows of the default weight of one (note that the w micro-flows share the same path). We refer to this approach as the *striped approach*. The connection, in this case, would theoretically receive w times the bandwidth enjoyed by a default connection *provided the aggregate bandwidth can be achieved*. The second approach involves using just one micro-flow, but adapting the congestion control to achieve a behavior that is equivalent to the aggregate behavior of w default micro-flows. The former approach has thus far been considered to be a non-viable solution because of the problems in effectively aggregating the bandwidths offered by the micro-flows [10]. On the other hand, related work that adopts the latter approach either does not scale beyond nominal weights (e.g. 10 in the case of mulTCP [4]), or relies on unrealistic assumptions (e.g. TCP-LASD [10] assumes accurate loss profiling), thus limiting the applicability.

4.2.1. Applicability

A clear application of pTCP is the striped approach where w micro-flows are used. Because of pTCP's unique ability to aggregate bandwidths across micro-flows that have widely differing and fluctuating bandwidths, the effective bandwidth perceived by the connection will indeed be w times that of a default connection. Two key issues of the striped approach to achieving rate differentiation are: (a) *Widely varying instantaneous bandwidths across the micro-flows*. Although the average throughput enjoyed by each individual micro-flow will approximately be the same, the instantaneous throughputs can be very different. (b) *Highly fluctuating bandwidths*. The bandwidth enjoyed by a micro-flow can by itself be fluctuating, especially when the number of micro-flows increases (for larger weights). pTCP handles the first issue through its delayed binding mechanism, and addresses the second issue by its packet re-striping strategy.

4.2.2. Performance

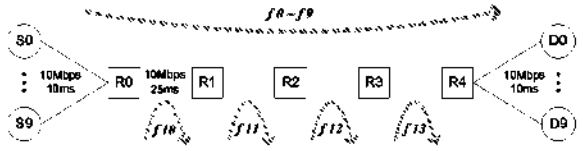
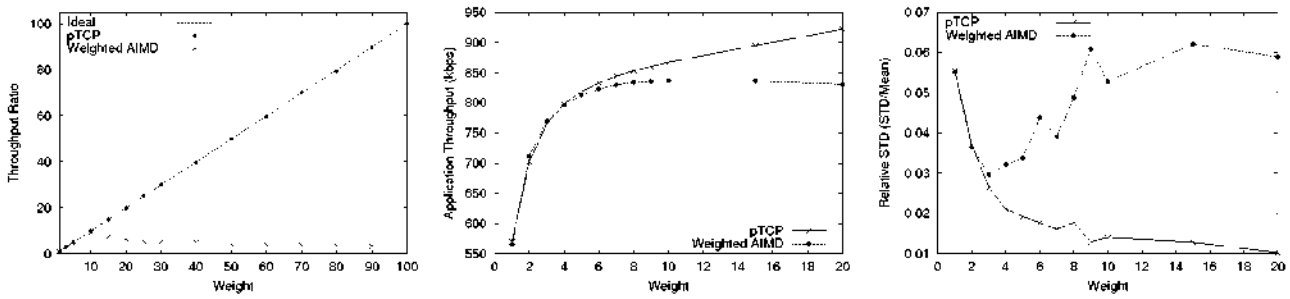


Figure 5. Network Topology for Weighted Rate Differentiation

Figure 5 shows the network topology used for evaluating the weighted rate differentiation instantiation of pTCP. It consists of 10 TCP flows with each having a 240ms round-trip time in a multi-link bottleneck environment. All backbone routers use RED queues with buffer size approximately equal to the bandwidth-delay product of the bottleneck link. The configuration roughly emulates the multi-

RED HAT 1010
RED HAT V. CAS



(a) Throughput Ratio (1 pTCP) (b) Throughput Mean (10 pTCPs) (c) Throughput Unfairness (10 pTCPs)

Figure 6. pTCP Performance for Weighted Rate Differentiation

plexing of a large number of flows at the backbone routers where each packet experiences approximately equal drop probability. We use two scenarios to compare the performance of pTCP and WAIMD (Weighted AIMD) [4] in providing service differentiation to end-to-end flows: (a) 1 pTCP/WAIMD flow with weight w and the other 9 regular TCP flows with weight 1, and (b) all 10 flows are pTCP/WAIMD flows with the same weight w. The first scenario is targeted toward evaluating the throughput scalability of pTCP with increasing weights, and the second scenario evaluates pTCP's fairness properties in this context. We also introduce 4 short TCP flows with 90ms round-trip time as the background traffic. We use TCP-SACK for all TCP flows and measure the throughput delivered to the application.

Figure 6(a) shows the simulation results for the first scenario where the weight of the pTCP/WAIMD flow increases from 1 to 100, while the other 9 flows are regular TCP flows with a weight of 1. It is clear that when using WAIMD the throughput of the weighted flow does not scale beyond 10. On the other hand, the performance of pTCP closely follows the ideal curve even when the weight is 100. The simulation result thus shows the effectiveness of pTCP in achieving scalable weighted service differentiation without relying on accurate loss measurement as in [10].

In Figure 6(b) and Figure 6(c) we show the performance of pTCP in the second scenario where all of the 10 flows are weighted flows with the same weight. The background TCP flows still have a weight of 1. As the weight of the 10 flows increases, the background TCP flows yield more bandwidth causing the throughput of the weighted flow to increase. As shown in the figure, however when WAIMD is used the throughput of the weighted flows saturates at $w - 8$ and starts decreasing afterward. This is a fundamental problem with the WAIMD scheme where each weighted TCP flow becomes sensitive to packet losses and its performance deteriorates as w increases. Moreover, although ideally all 10 flows should enjoy the same throughput we show in Figure 6(c) the throughput unfairness (standard deviation normalized to the mean) observed among the 10 flows. Because pTCP is able to effectively aggregate the throughput of the component micro-flows, as weight increases any unfairness introduced in the router (e.g. due to unequal packet drop probability) on individual micro-flows is reduced. On the other hand, since WAIMD flows become sensitive to packet losses at large weights, the unfairness in fact increases as the weight increases. Therefore, we show that using pTCP for end-to-end service differentiation results in better scalability and fairness than that demonstrated in related work [4, 10].

4.3. Application Layer Switching

Several Internet performance studies have identified two characteristics of the Internet: (a) Predominantly, there exists a path different from the default one provided by the underlying routing infrastructure that can provide better service (in terms of bandwidth or delay) [11]. (b) The Internet infrastructure remains under-utilized in pockets, while it is overloaded in other areas [3]. A combination of the above two observations has led to research in exploiting such unbalanced use of the Internet and providing better quality of service (QoS) to connections. Since it is desirable for any Internet solution to be easily deployable, an approach that is based on an overlay network constructed through purely application layer switching holds great promise.

The idea is to construct an overlay network without changing the existing Internet infrastructure. The overlay network will comprise of only Internet end-hosts that will act as application layer switches.[3] The participating end-hosts also double up as QoS monitors, measuring the QoS observed through the "links" on the overlay network. Traffic on the overlay network can be routed through multiple "hops" based on the QoS observed on the different possible routes. Thus, although the default path from say a host in

[3]Note that although conceptually we refer to such a technique as application layer switching, a practical implementation can be more efficient (say using IPv4 source routing).

RED HAT 1010
RED HAT V. CAS

Chicago to a host in Washington might exist, traffic between the two hosts might be routed through another host in San Francisco if the alternate path provides better QoS.

A critical obstacle in realizing such an approach lies in the QoS monitoring at the participating end-hosts. Most of the studies that have identified the existence of better alternate paths have done so using invasive probing techniques, physically injecting data into the network paths. Such an approach as a long-term solution would however be quite undesirable since it consumes valuable network resources. Inferring available QoS from existing connections between the participating end-hosts might be a tall order as the traffic on a "link" between two end-hosts might be composed of traffic from non-participating end-hosts as well. Invasive probing in intervals can render the routing decisions potentially stale, and the overheads of probing still remain.

In this context, we consider a solution wherein k different paths on the overlay network (between the source and the destination) that are coarsely found to exhibit the best QoS are chosen, and the end-to-end connection is striped across the k paths. As long as the aggregate bandwidth can be delivered to the connection, such an approach would entail only finding the k best paths as opposed to the best path, and hence is more feasible. Even infrequent invasive probing can establish the k best paths on the overlay network. Such an approach would allow the striped connection to always enjoy the services along the best path as long as it lies within the k chosen paths.

Note that in order to alleviate unfairness issues, all connections on the overlay network can be striped over k paths. As k increases, not only should the throughput and fairness properties experienced by connections increase, but the utilization of network resources should also improve considerably. Even when the network is saturated, a larger k should result in better global fairness between the connections in the network.

4.3.1. Applicability

The applicability of pTCP in this domain is quite evident. Striping of connections is advantageous as long as the aggregate bandwidths can be achieved. The different paths available for a striped connection can exhibit both long-term and short-term bandwidth and delay fluctuations. pTCP is designed to exactly operate in such scenarios and provide the striped connection with the aggregate bandwidth.

4.3.2. Performance

Figure 7 shows the network topology used to simulate striping using application layer switching in an overlay network. We assume existence of 10 different paths between adjacent backbone routers. All of the 40 backbone links

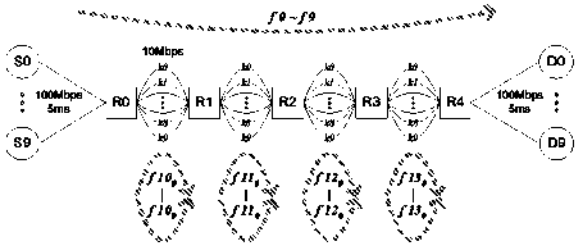


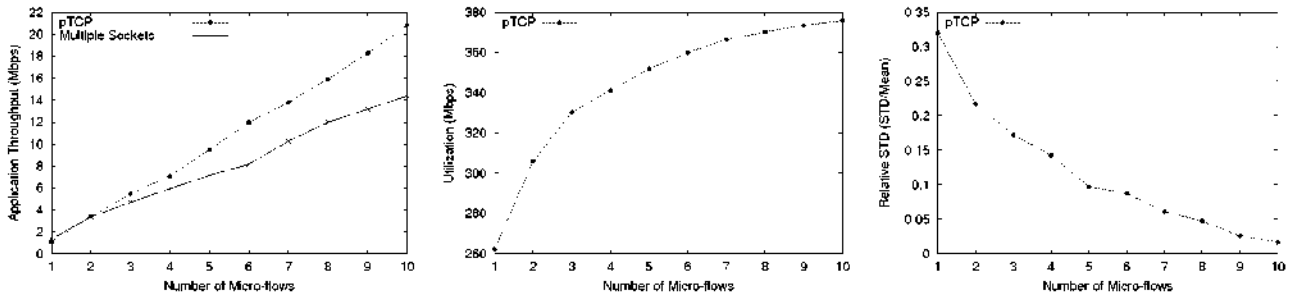
Figure 7. Network Topology for Application Layer Switching

are 10Mbps links with propagation delay randomly chosen from 5ms to 30ms. We use a UDP flow on each of the backbone link (total 40 UDP flows) to emulate the background traffic. The data rates of the UDP flows are randomly chosen from 3Mbps to 7Mbps and is changed after every 10-second period. This configuration thus represents an average of 50% utilization on the backbone links. Note that in the *ns-2* simulation, because different links originating from the same node are in fact independent of each other (they do not share the same queue), using such network topology is not limited to the scenario when all routers have 10 fan-in and fan-out links. Instead, it emulates a more generic topology where there are 10000 different paths (each may be with a different round-trip time and traffic loads but not necessarily disjoint) between $R0$ and $R4$. We use 10 TCP flows and measure their throughput at the respective receiving applications. We use two scenarios to show the performance of pTCP in such an environment: (a) 1 pTCP flow with p micro-flows and 9 regular TCP flows, and (b) 10 pTCP flows each with p micro-flows. Each micro-flow takes on a path randomly chosen from the 10000 end-to-end paths existing in the backbone network.

Figure 8(a) shows the throughput enjoyed by the single pTCP flow when all the other 9 flows are regular TCP flows. As the the number of micro-flows composing the pTCP flow increases, the throughput for the pTCP flow increases by utilizing the unused resources in the network.

In Figure 8(b) and Figure 8(c) we show the simulation results when all the 10 flows are pTCP flows with the same number of micro-flows (each of them randomly takes on an end-to-end path in the backbone network). Figure 8(b) shows the network utilization that is measured by summing the throughput of background UDP traffic and the end-to-end throughput of all pTCP flows multiplied by the number of links (which is 4), and Figure 8(c) shows throughput unfairness among the 10 flows using the normalized standard deviation. *It is clear that not only does the network utilization increase (the upper bound is 400Mbps) as each pTCP uses more micro-flows, but the unfairness (due to micro-flows traversing different bottleneck links) also reduces.*

RED HAT 1010
RED HAT V. CAS



(a) Application Throughput (1 pTCP) (b) Network Utilization (10 pTCPs) (c) Network Unfairness (10 pTCPs)

Figure 8. pTCP Performance for Application Layer Switching

5. Related Work and Conclusions

Striping is a technique used for the transparent aggregation of multiple resources to obtain higher performance [16]. It has been typically used at the link layer to aggregate bandwidths of different links with similar characteristics [1, 5]. Recently, several application layer striping approaches that open multiple TCP sockets in parallel to achieve higher throughput have also been proposed [2, 6, 8, 12]. However, these approaches deal with increasing application throughput by using multiple TCP connections *through the same physical path*. If such approaches are used for striped connections over different paths, the problems identified in Section 2 would render them ineffective.

In this paper, we propose a transport protocol called pTCP that is suited for supporting striped connections. pTCP achieves aggregate bandwidths for striped connections even in the presence of disparities among the different paths, and fluctuations in path characteristics. We show the applicability of pTCP in three different application domains: (a) bandwidth aggregation on multi-homed mobile hosts, (b) end-to-end service differentiation, and (c) striping on overlay networks based on application layer switching. pTCP demonstrates considerably improved performance in each of the three settings when compared to existing approaches.

References

[1] H. Adiseshu, G. Parulkar, and G. Varghese. A reliable and scalable striping protocol. In *Proceedings of ACM SIGCOMM*, Palo Alto, CA USA, Aug. 1996.

[2] M. Allman, H. Kruse, and S. Ostermann. An application-level solution to TCP's satellite inefficiencies. In *Proceedings of WOSBIS*, Rye, NY USA, Nov. 1996.

[3] CAIDA. Internet measurement and data analysis. http://www.caida.org/analysis/performance.

[4] J. Crowcroft and P. Oechslin. Differentiated end-to-end Internet services using a weighted proportional fair sharing TCP. *ACM Computer Communication Review*, 28(3):53–69, July 1998.

[5] J. Duncanson. Inverse multiplexing. *IEEE Communications Magazine*, 32(4):34–41, Apr. 1994.

[6] T. Hacker, B. Athey, and B. Noble. The end-to-end performance effects of parallel TCP sockets on a lossy wide-area network. In *Proceedings of IPDPS*, Fort Lauderdale, FL USA, Apr. 2002.

[7] H.-Y. Hsieh and R. Sivakumar. A transport layer approach fo achieving aggregate bandwidths on multi-homed mobile hosts. In *Proceedings of ACM MOBICOM*, Atlanta, GA USA, Sept. 2002.

[8] J. Lee, D. Gunter, B. Tierney, B. Allcock, J. Bester, J. Bresnahan, and S. Tuecke. Applied techniques for high bandwidth data transfers across wide area networks. In *Proceedings of CHEP*, Beijing, China, Sept. 2001.

[9] L. Magalhaes and R. Kravets. Transport level mechanisms for bandwidth aggregation on mobile hosts. In *Proceedings of IEEE ICNP*, Riverside, CA USA, Nov. 2001.

[10] T. Nandagopal, K.-W. Lee, J.-R. Li, and V. Bharghavan. Scalable service differentiation using purely end-to-end mechanisms: Features and limitations. In *Proceedings of IWQoS*, Pittsburgh, PA USA, June 2000.

[11] S. Savage, A. Collins, E. Hoffman, J. Snell, and T. Anderson. The end-to-end effects of Internet path selection. In *Proceedings of ACM SIGCOMM*, Cambridge, MA USA, Sept. 1999.

[12] H. Sivakumar, S. Bailey, and R. Grossman. PSockets: The case for application-level network striping for data intensive applications using high speed wide area networks. In *Proceedings of Supercomputing (SC)*, Dallas, TX USA, Nov. 2000.

[13] A. Snoeren. Adaptive inverse multiplexing for wide-area wireless networks. In *Proceedings of IEEE GLOBECOM*, Rio de Janeireo, Brazil, Dec. 1999.

[14] M. Stemm and R. Katz. Vertical handoffs in wireless overlay networks. *Mobile Networks and Applications*, 3(4):335–350, 1998.

[15] The Network Simulator *ns-2*. http://www.isi.edu/nsnam/ns.

[16] B. Traw and J. Smith. Striping within the network subsystem. *IEEE Network*, 9(4):22–32, July 1995.

[17] G. R. Wright and W. R. Stevens. *TCP/IP Illustrated, Volume 2: The Implementation*. Addison-Wesley Publishing Company, Reading, MA USA, Jan. 1995.

RED HAT 1010
RED HAT V. CAS

Electronic Acknowledgement Receipt

EFS ID:	20613066
Application Number:	14512414
International Application Number:	
Confirmation Number:	9530
Title of Invention:	MULTIPATH COMMUNICATION DEVICES AND METHODS
First Named Inventor/Applicant Name:	Eric M. Delangis
Customer Number:	64828
Filer:	Harold C. McGurk
Filer Authorized By:	
Attorney Docket Number:	ED 2000
Receipt Date:	05-NOV-2014
Filing Date:	11-OCT-2014
Time Stamp:	14:06:31
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Information Disclosure Statement (IDS) Form (SB08)	14512414_IDS_11-5-2014.pdf	612224 fbce5 1da89e8dc9f888c9 1e8b 78cffb495ce93 6ca	no	4

Warnings:	
Information:	

RED HAT 1010
RED HAT V. CAS

A U.S. Patent Number Citation or a U.S. Publication Number Citation is required in the Information Disclosure Statement (IDS) form for autoloading of data into USPTO systems. You may remove the form to add the required data in order to correct the Informational Message if you are citing U.S. References. If you chose not to include U.S. References, the image of the form will be processed and be made available within the Image File Wrapper (IFW) system. However, no data will be extracted from this form. Any additional data such as Foreign Patent Documents or Non Patent Literature will be manually reviewed and keyed into USPTO systems.

2	Transmittal Letter	14512414_IDS_Transmittal_Letter.pdf	72478 021303569401f7795a4d1131388f0a7220e5567	no	1
Warnings:					
Information:					
3	Non Patent Literature	Fisk.pdf	637190 2fa3b382eb678128a330c85d99670c9edad98eba	no	3
Warnings:					
Information:					
4	Non Patent Literature	Hsieh.pdf	191193 a09564a58e769477153d3a0c0c6418868c11b3d0	no	10
Warnings:					
Information:					
		Total Files Size (in bytes):	1513085		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING OR 371(C) DATE	FIRST NAMED APPLICANT	ATTY. DOCKET NO./TITLE
14/512,414	10/11/2014	Eric M. Delangis	ED 2000

CONFIRMATION NO. 9530

64828
Harold McGurk
The Law Office of Clay McGurk
P.O. BOX 1488
Orange, CA 92856

PUBLICATION NOTICE

OC000000073103365

Title:MULTIPATH COMMUNICATION DEVICES AND METHODS

Publication No.US-2015-0032860-A1
Publication Date:01/29/2015

NOTICE OF PUBLICATION OF APPLICATION

The above-identified application will be electronically published as a patent application publication pursuant to 37 CFR 1.211, et seq. The patent application publication number and publication date are set forth above.

The publication may be accessed through the USPTO's publically available Searchable Databases via the Internet at www.uspto.gov. The direct link to access the publication is currently http://www.uspto.gov/patft/.

The publication process established by the Office does not provide for mailing a copy of the publication to applicant. A copy of the publication may be obtained from the Office upon payment of the appropriate fee set forth in 37 CFR 1.19(a)(1). Orders for copies of patent application publications are handled by the USPTO's Office of Public Records. The Office of Public Records can be reached by telephone at (703) 308-9726 or (800) 972-6382, by facsimile at (703) 305-8759, by mail addressed to the United States Patent and Trademark Office, Office of Public Records, Alexandria, VA 22313-1450 or via the Internet.

In addition, information on the status of the application, including the mailing date of Office actions and the dates of receipt of correspondence filed in the Office, may also be accessed via the Internet through the Patent Electronic Business Center at www.uspto.gov using the public side of the Patent Application Information and Retrieval (PAIR) system. The direct link to access this status information is currently http://pair.uspto.gov/. Prior to publication, such status information is confidential and may only be obtained by applicant using the private side of PAIR.

Further assistance in electronically accessing the publication, or about PAIR, is available by calling the Patent Electronic Business Center at 1-866-217-9197.

Office of Data Managment, Application Assistance Unit (571) 272-4000, or (571) 272-4200, or 1-888-786-0101

To: CLAYMCGURK@GMAIL.COM,,
From: PAIR_eOfficeAction@uspto.gov
Cc: PAIR_eOfficeAction@uspto.gov
Subject: Private PAIR Correspondence Notification for Customer Number 64828

Jan 30, 2015 05:39:01 AM

Dear PAIR Customer:

Harold McGurk
The Law Office of Clay McGurk
P.O. BOX 1488
Orange, CA 92856
UNITED STATES

The following USPTO patent application(s) associated with your Customer Number, 64828 , have new outgoing correspondence. This correspondence is now available for viewing in Private PAIR.

The official date of notification of the outgoing correspondence will be indicated on the form PTOL-90 accompanying the correspondence.

Disclaimer:
The list of documents shown below is provided as a courtesy and is not part of the official file wrapper. The content of the images shown in PAIR is the official record.

Application	Document	Mailroom Date	Attorney Docket No.
14512414	NTC.PUB	01/29/2015	ED 2000

To view your correspondence online or update your email addresses, please visit us anytime at https://sportal.uspto.gov/secure/myportal/privatepair.

If you have any questions, please email the Electronic Business Center (EBC) at EBC@uspto.gov with 'e-Office Action' on the subject line or call 1-866-217-9197 during the following hours:

Monday - Friday 6:00 a.m. to 12:00 a.m.

Thank you for prompt attention to this notice,

UNITED STATES PATENT AND TRADEMARK OFFICE

PATENT APPLICATION INFORMATION RETRIEVAL SYSTEM

RED HAT 1010
RED HAT V. CAS

Doc code: IDS
Doc description: Information Disclosure Statement (IDS) Filed

PTO/SB/08a (01-10)
Approved for use through 07/31/2012. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (**Not for submission under 37 CFR 1.99)**		
	Application Number	14512414
	Filing Date	2014-10-11
	First Named Inventor	Eric Delangis
	Art Unit	2652
	Examiner Name	
	Attorney Docket Number	ED 2000

U.S.PATENTS Remove

Examiner Initial*	Cite No	Patent Number	Kind Code[1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1					

If you wish to add additional U.S. Patent citation information please click the Add button. Add

U.S.PATENT APPLICATION PUBLICATIONS Remove

Examiner Initial*	Cite No	Publication Number	Kind Code[1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1					

If you wish to add additional U.S. Published Application citation information please click the Add button. Add

FOREIGN PATENT DOCUMENTS Remove

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2] i	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	T[5]
	1							☐

If you wish to add additional Foreign Patent Document citation information please click the Add button Add

NON-PATENT LITERATURE DOCUMENTS Remove

Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc), date, pages(s), volume-issue number(s), publisher, city and/or country where published.	T[5]

EFS Web 2.1.17

RED HAT 1010
RED HAT V. CAS

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	14512414
	Filing Date	2014-10-11
	First Named Inventor	Eric Delangis
	Art Unit	2652
	Examiner Name	
	Attorney Docket Number	ED 2000

	1	"End-to-End Approach to Host Mobility", by Alex C. Snoeren and Hari Balakrishnan, MIT Laboratory for Computer Science, MobiCom (2000).	☐
	2	"Multi-homed TCP", by Huitema, C. (Internet Draft) IETF, (1995).	☐
	3	"TCP extensions for high performance", by Jacobson V., Braden R., and Borman D., RFC 1323, IETF, (1992).	☐
	4	"A New Multicasting-based Architecture for Internet Host Mobility", by Mysore, J., and Bharghavan, V., Proc. ACM/ IEEE Mobicom '97 (1997).	☐
	5	"Mobile IPv4 Challenge/Response Extensions", by Perkins, C., and Calhoun, P., RFC 3012, IETF, (2000).	☐
	6	"Mobility Support in IPv6", by Perkins, C., and Johnson, D, Proc. ACM/IEEE Mobicom '96 (1996).	☐
	7	"Route Optimization in Mobile IP", by Perkins C., and Johnson, D., Cluster Computing 1, (1998).	☐
	8	"IP Mobility Support", by Perkins, C., RFC 2002, IETF, (1996).	☐
	9	"Transmission Control Protocol", RFC 793, IETF (1981).	☐

If you wish to add additional non-patent literature document citation information please click the Add button Add

EXAMINER SIGNATURE

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

EFS Web 2.1.17

RED HAT 1010
RED HAT V. CAS

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	14512414
	Filing Date	2014-10-11
	First Named Inventor	Eric Delangis
	Art Unit	2652
	Examiner Name	
	Attorney Docket Number	ED 2000

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	14512414
	Filing Date	2014-10-11
	First Named Inventor	Eric Delangis
	Art Unit	2652
	Examiner Name	
	Attorney Docket Number	ED 2000

CERTIFICATION STATEMENT

Please see 37 CFR 1.97 and 1.98 to make the appropriate selection(s):

☐ That each item of information contained in the information disclosure statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(1).

OR

☐ That no item of information contained in the information disclosure statement was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the person signing the certification after making reasonable inquiry, no item of information contained in the information disclosure statement was known to any individual designated in 37 CFR 1.56(c) more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(2).

☐ See attached certification statement.

☐ The fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☐ A certification statement is not submitted herewith.

SIGNATURE

A signature of the applicant or representative is required in accordance with CFR 1.33, 10.18. Please see CFR 1.4(d) for the form of the signature.

Signature	/H C McGurk IV/	Date (YYYY-MM-DD)	2015-07-02
Name/Print	H C McGurk	Registration Number	34964

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 1 hour to complete, including gathering, preparing and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these record s.

2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.

3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.

4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).

5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.

6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).

7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.

8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.

9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

Second Annual International Conference
Mobile Computing and Networking
(MobiCom'96), Rye, New York,
USA,
November 1996.

Mobility Support in IPv6

Charles E. Perkins
&
David B. Johnson

Presented By:
Ajay Sharma.

WPI



RED HAT 1010
RED HAT V. CAS

About The Author

Charles E. Perkins: Research Fellow at Nokia Research Center investigating mobile wireless networking and dynamic configuration protocols. He is the editor for several ACM and IEEE journals for areas relating to wireless networking. Charles has served on the Internet Architecture Board (IAB) and on various committees for the National Research Council. He has published a number of papers and award-winning articles in the areas of mobile networking, resource discovery, and automatic configuration for mobile computers.

David B. Johnson: Associate Professor of Computer Science and Electrical and Computer Engineering at Rice University . He was a principal designers of the IETF Mobile IP protocol for IPv4 and primary designer of Mobile IP for IPv6. Currently an Executive Committee member and the Treasurer for SIGMOBILE, also a member of the Editorial Board for IEEE/ACM Transactions on Wireless Networks.

Outline

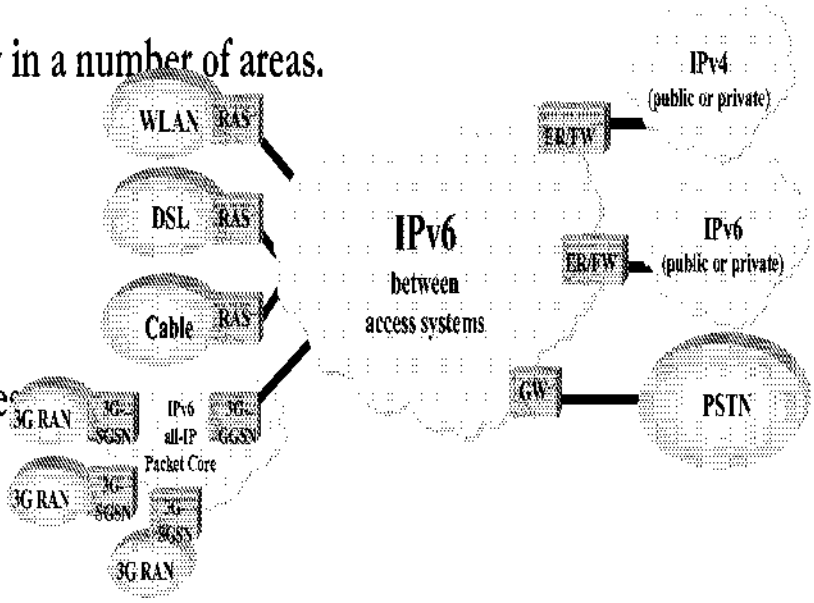
1. Why Mobile IPv6
2. Benefits of Mobile IPv6
3. What is IPv6?
4. Address Architecture of IPv6.
5. Mobile IPv6 Terminology.
6. Mobile IPv6 Mechanism.
7. Errors Handling
8. Security Handling
9. Summery
10. Q & A



RED HAT 1010
RED HAT V. CAS

Why Mobile IPv6? -- Propellant factors.

- Huge growth of mobile Internet terminals will exhaust IPv4 address space
  - All wireless terminals will have WAP and GPRS
    - IPv6 brings enough IP addresses
- Ease of scalability
  - Supporting billions of new devices and huge amounts of new bandwidth
  - Simplified, cost-efficient architecture without NATs , Proxies, ALGs,...
- Always-on connection establishes a variety of new services.
  - Push, location-based, etc.
- Integrated Security
- Efficiency: IPv6 improves efficiency in a number of areas.
  - Routing, Broadcast handling
- Quality of Service improvements
  - Fragmentation, Flows
- Mobility Across Access Technologies



WPI

RED HAT 1010
RED HAT V. CAS

Requirements for Mobility in Internet

Mobility	• Increasing number of users asks for Mobility Support in Internet
Transparency	• Mobility shall be transparent to all Protocol Layers above IP
Easy to use	• Mobility shall be as easy to handle as with Mobile Phones in GSM
Routing	• Mobility shall be compatible to all Routing Protocols and shall optimize routes
Security	• Mobility shall not decrease security in Internet

IPv6 features relevant to Mobile IP

- Larger address space => Unique Global address for each device. ($6.65 \cdot 10^{23}$ addresses per m^2 of earth surface)
- Scalable => Run over multiple media i.e. Wireless-LAN, Ethernet, 3G
- Auto configuration capabilities=> Network Plug-and-Play.
- Fixed header format => Fewer fields (8 as compared to 12 in IPv4)
- Router headers => MIP updates are in extension headers. No header length anymore.
- Security extensions => Internet level Security in IPv6 Header.
- Anycast addresses => Special type of address in IPv6.
- Encapsulation =>IP-layer authentication & encryption possible.
- Quality of service and flow labels => efficient routing for real-time applications.
- Elimination of "triangle routing" for mobile IP
- All nodes can handle bindings.
- Small overhead for distributing bindings. Fixed header format
- option extension headers not parsed by intermediate routers anymore

Basic IPv6 Address Types

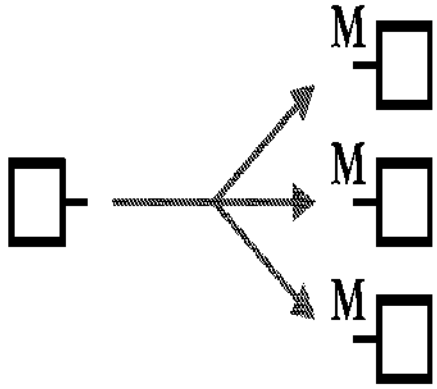
unicast:

for one-to-one communication



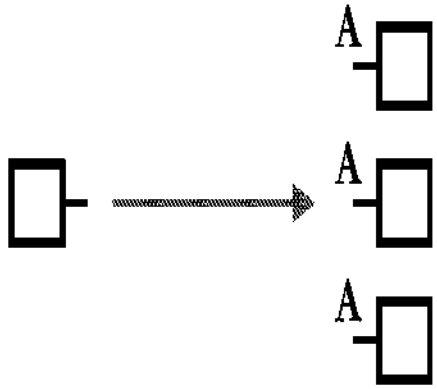
multicast:

for one-to-many communication



anycast:

for one-to-nearest communication



7

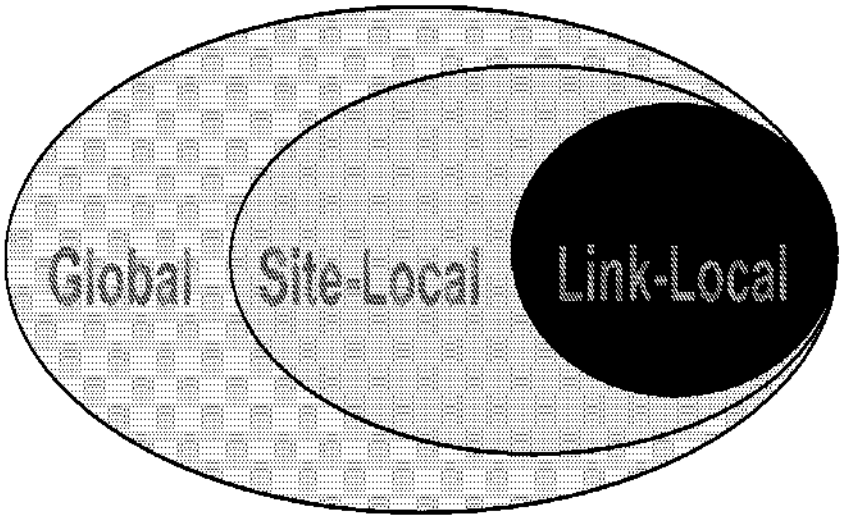
WPI



RED HAT 1010
RED HAT V. CAS

IPv6 - Addressing Model

- addresses are assigned to interfaces
  - No change from IPv4 Model
- interface 'expected' to have multiple addresses
- addresses have scope
  - Link Local
  - Site Local
  - Global



- addresses have lifetime
  - Valid and Preferred lifetime

WPI



RED HAT 1010
RED HAT V. CAS

Text Representation of IPv6 Address

"Preferred" form: 1080:0:FF:0:8:800:200C:417A

Compressed form: FF01:0:0:0:0:0:0:43 becomes FF01::43

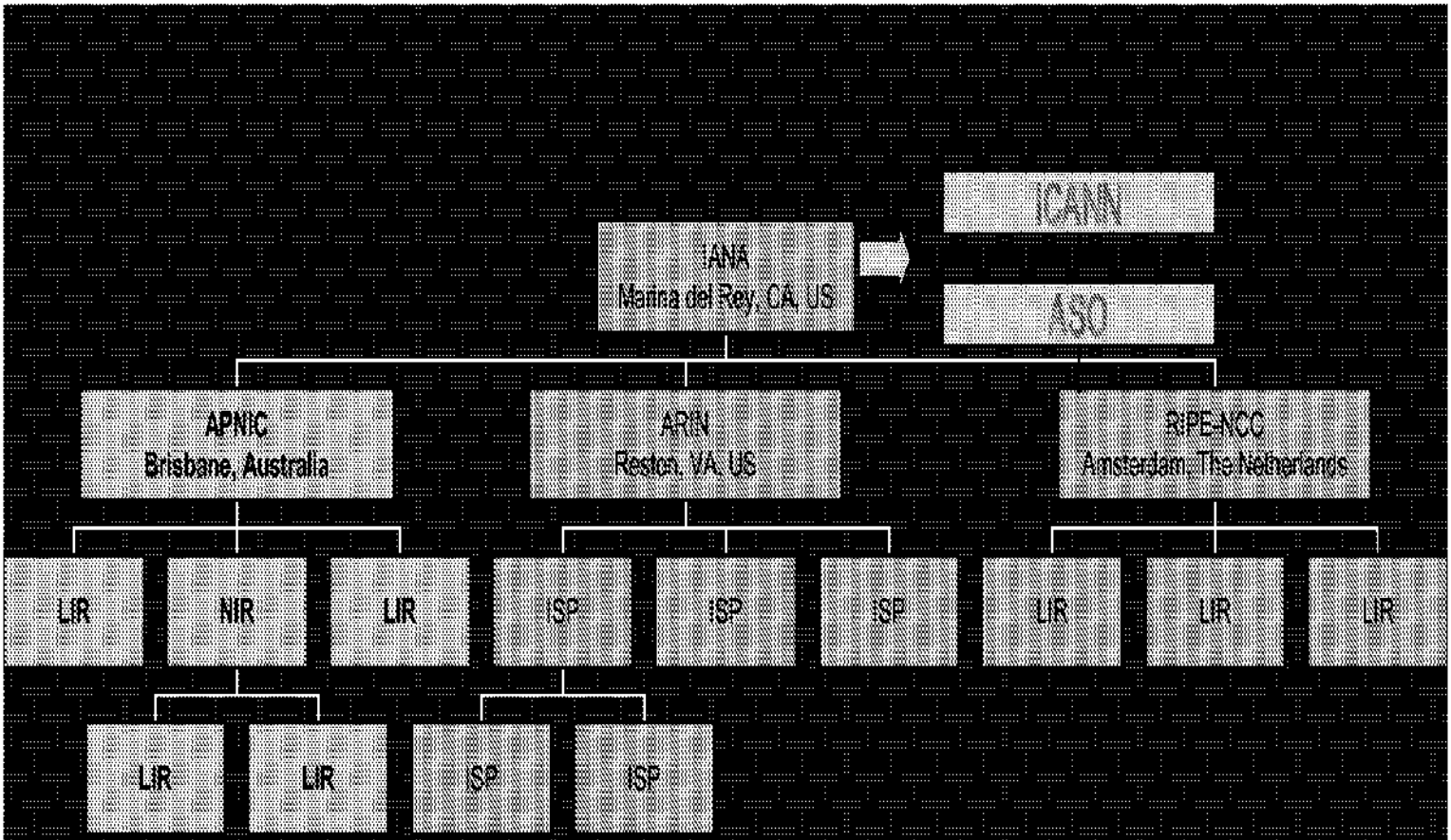
IPv4-compatible: 0:0:0:0:0:0:13.1.68.3 or ::13.1.68.3

There is no broadcast addresses, only multicast.

Loopback address is ::1



Internet Registry Hierarchy



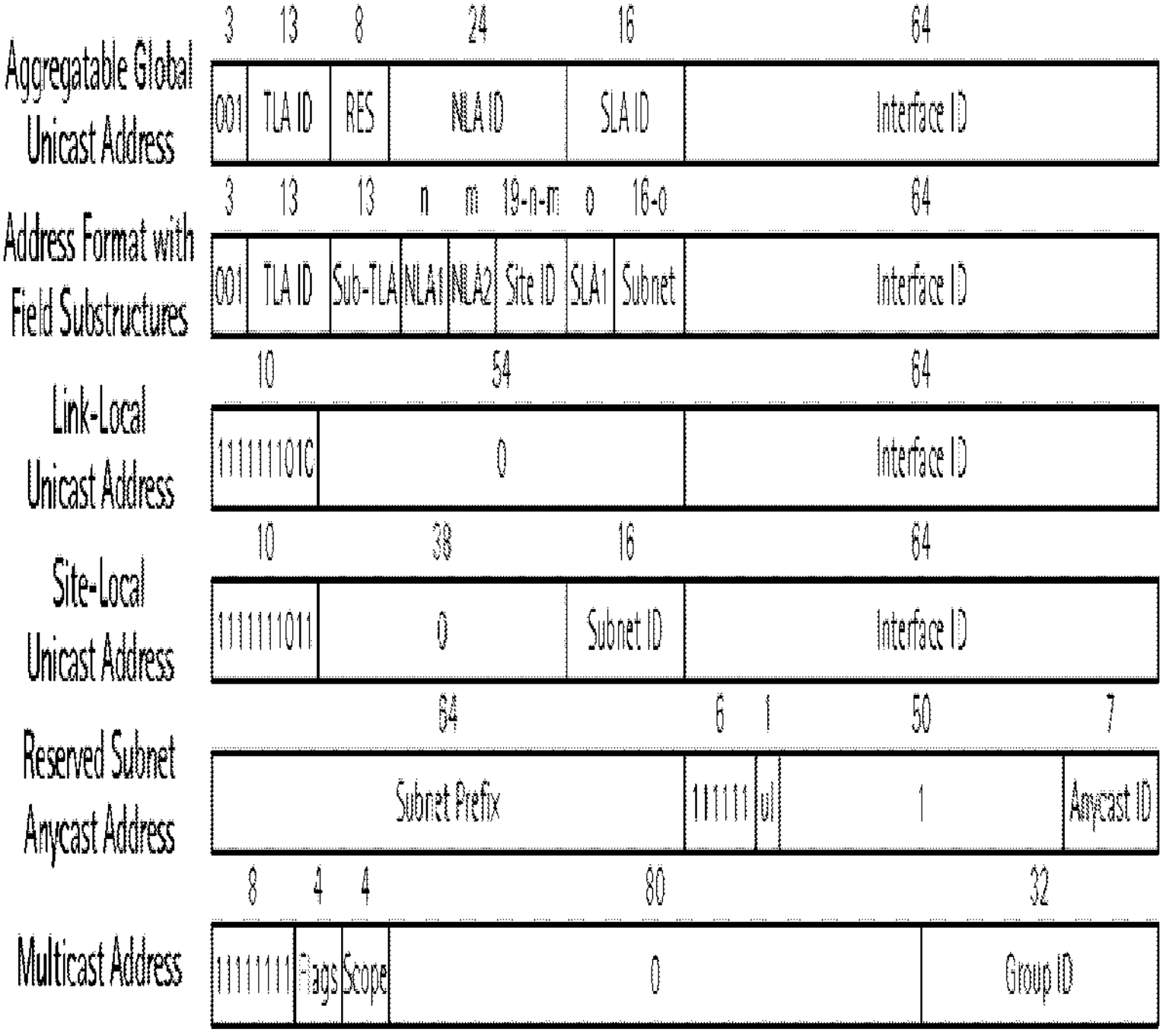
- **ICANN**: The Internet Corporation for Assigned Name and Number
- **ASO:** Address Supporting Organization.
- **IANA:** Internet Assigned Number Authority.
- **ARIN:** American Registry for Internet Number.
- **APNIC:** Asia Pacific Network Information Centre.
- **RIPE-NCC:** Reseaux IP Europeene.

WPI



RED HAT 1010
RED HAT V. CAS

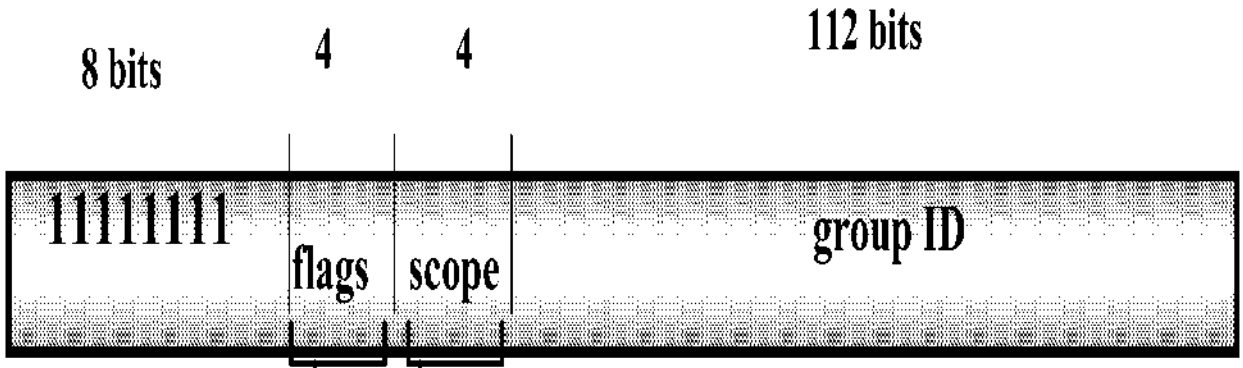
IPv6 Address Formats





RED HAT 1010
RED HAT V. CAS

Multicast address



0 reserved
1 node-local scope
2 link-local scope
5 site-local scope
8 organization-local scope
E global scope
F reserved

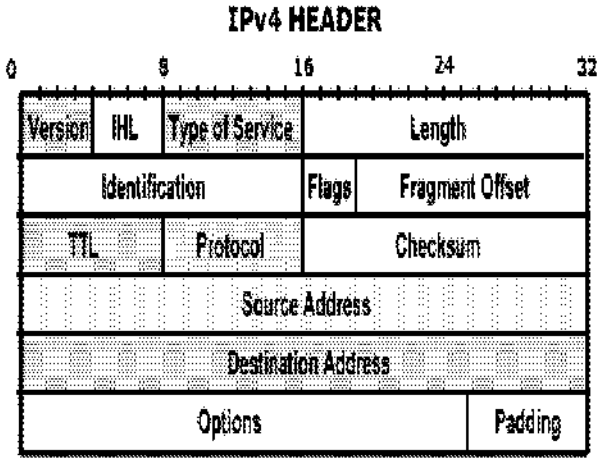
0000 Permanent address (by number authority)
0001 Transient address (can be established by appl.

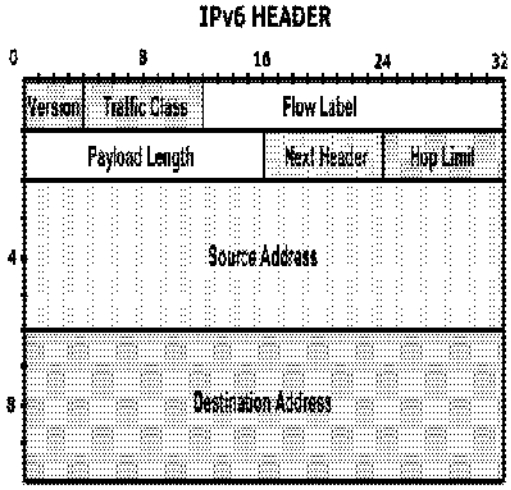
WPI



RED HAT 1010
RED HAT V. CAS

IPv4 vs. IPv6 Header



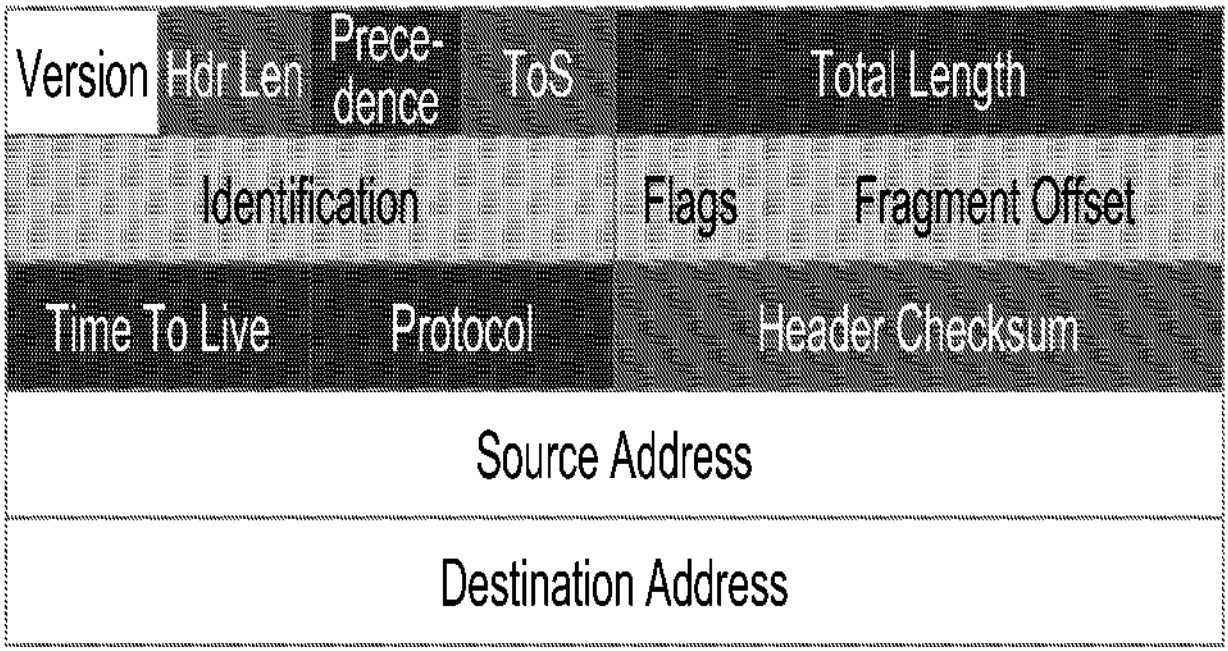


- 14 fields, at least 20 octets
- 32 bit addresses
- fragmented packet processing at every hop
- header checksum recalculation at every hop
- variable Options field for extra processing information

- 8 fields, fixed 40 octet size
- 128 bit addresses
- fragmentation only in src and dst endpoint, or lower layer
- no checksums
- new 20 bit flow label field
- options in Extension Headers

Changes in IPv4 Header

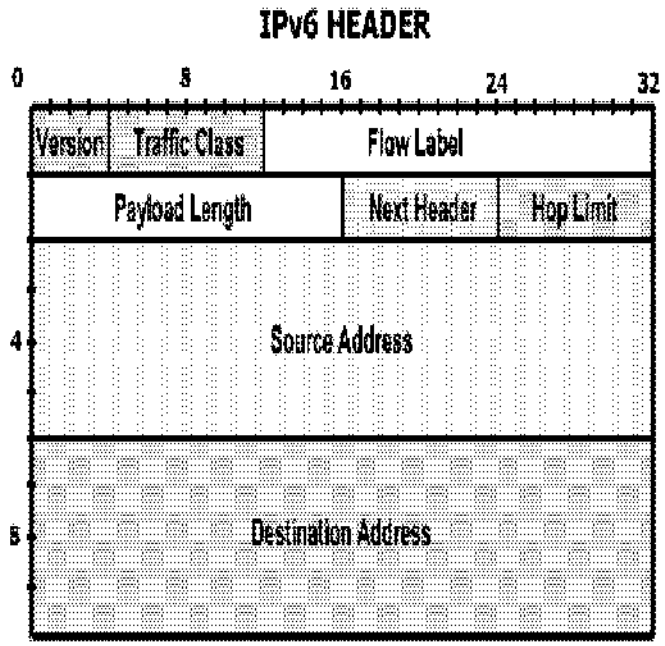
- 20 bytes
- 13 fields
- removed
- moved to extension headers
- renamed
  - –precedence → class
  - –total length → payload length
  - –time to live → hop limit
  - –protocol → next header





RED HAT 1010
RED HAT V. CAS

IPv6 Header Simplifications



Simplifications

Fixed format headers

no options -> no need for header length

options expressed as Extension headers

No header checksum

reduce cost of header processing, no checksum updates at each router

minimal risk as encapsulation of media access protocols (e.g...., Ethernet, PPP) have checksum

No segmentation

hosts should use path MTU discovery

otherwise use the minimum MTU (536 bytes)

- Base header is fixed size - 40 octets
  - NEXT HEADER field in base header defines type of header
  - Appears at end of fixed-size base header
- Some extensions headers are variable sized
  - NEXT HEADER field in extension header defines type
  - HEADER LEN field gives size of extension header

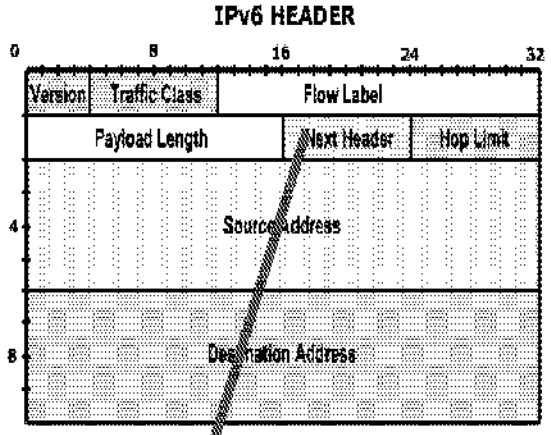
WPI



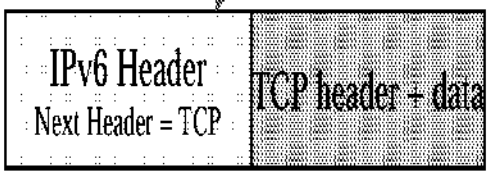
RED HAT 1010
RED HAT V. CAS

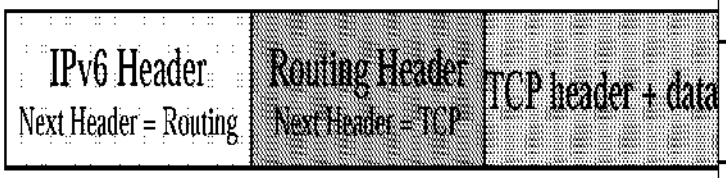
Extension Header

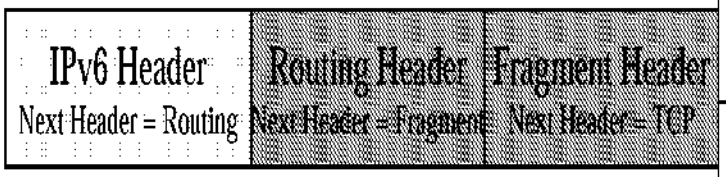
Store optional internet-layer information [Placed between IPv6 header and upper-layer header]

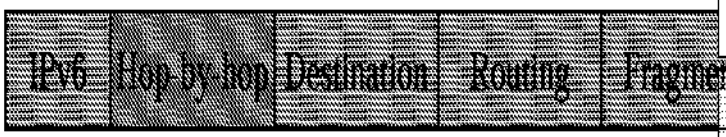


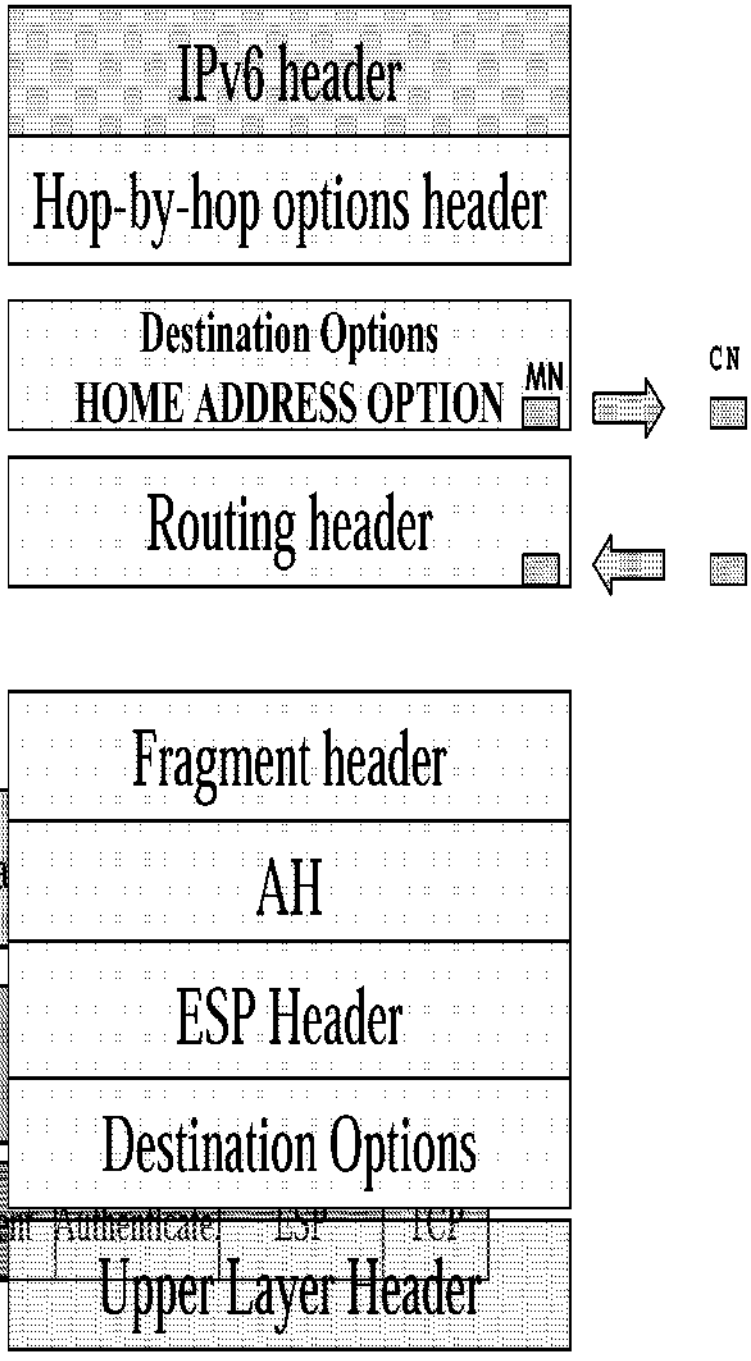
The Next Header field points to an extension Header









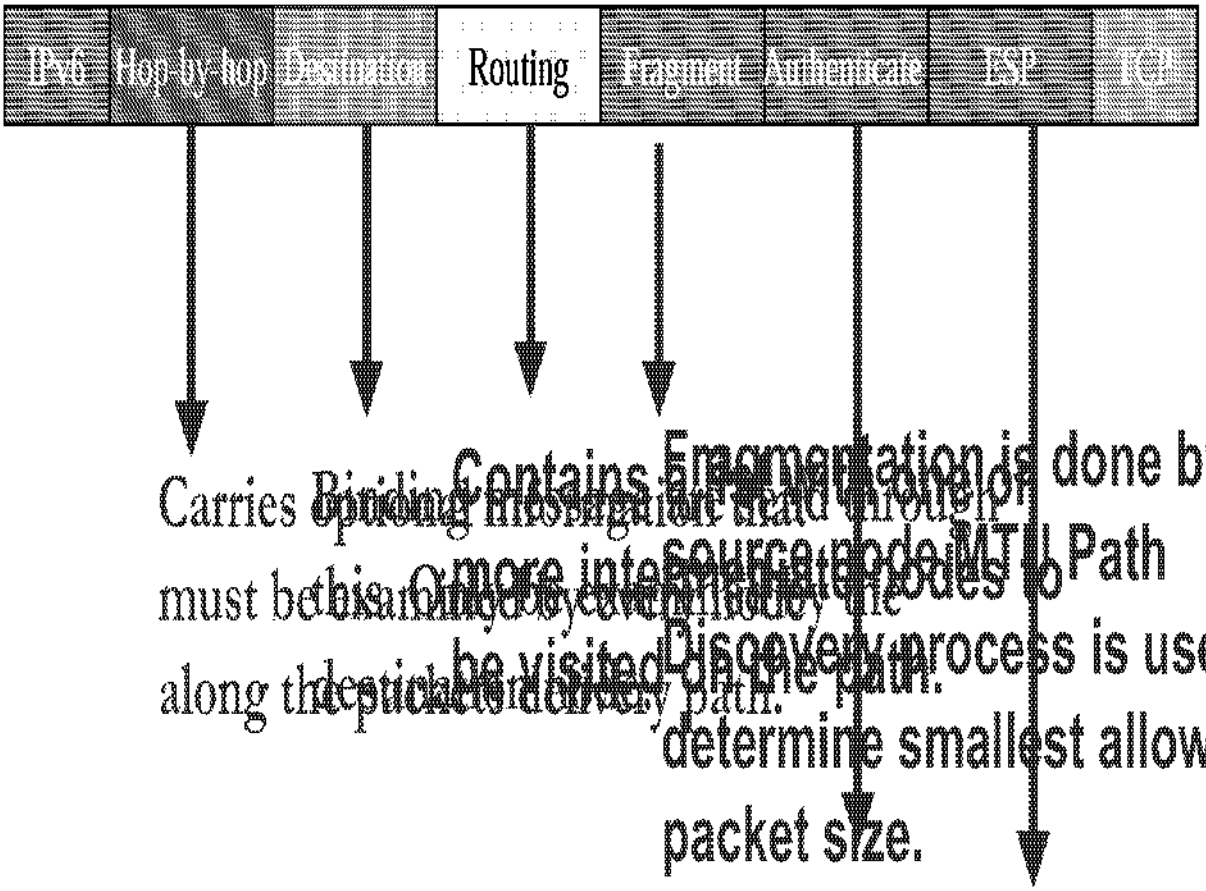


WPI



RED HAT 1010
RED HAT V. CAS

Extension Header



WPI



RED HAT 1010
RED HAT V. CAS

Terms used in Mobile IPv6

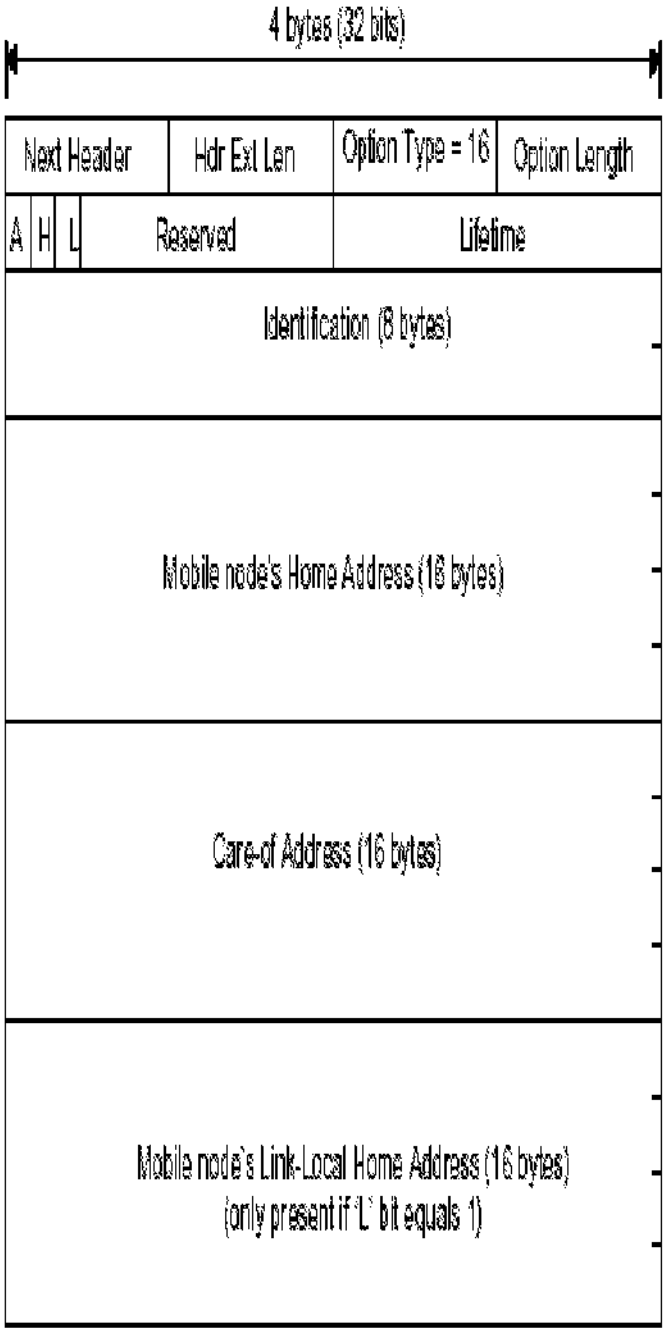
Mobile Node	Node, which can change its access point to the Internet while still being reachable under its Home Address.
Home Address	Static IP Address of the Mobile Node valid at its home network.
C/o-Address	Temporary IP Address of the Mobile Node valid at the actually visited network of the Mobile Node (c/o = care-of).
Binding	Association of the Home Address with the c/o-Address.
Home Agent	Router located at the Mobile Node's home network used by the Mobile Node for registering its c/o-Address.
Binding Cache	Cache for received Bindings.

WPI



RED HAT 1010
RED HAT V. CAS

Binding Update Option Header Format



A Bit : Indicates whether receiver should reply or not with Binding Acknowledgement.

H Bit: Use when mobile node wants the receiving node to act a Home Agent.

L Bit: Set if the mobile node want to receive packet destined to its link-local address.

Lifetime: Lease time for the address.

Identification Field: Counter is use to insure Binding Updates are order-wise. Counter increment for each new BU (not for retransmission) .

Care-of Address: current address of MN. When care-of address = Home address. Destination Cache entries should be deleted.



RED HAT 1010
RED HAT V. CAS

Server-less Autoconfiguration ("Plug-n-Play")

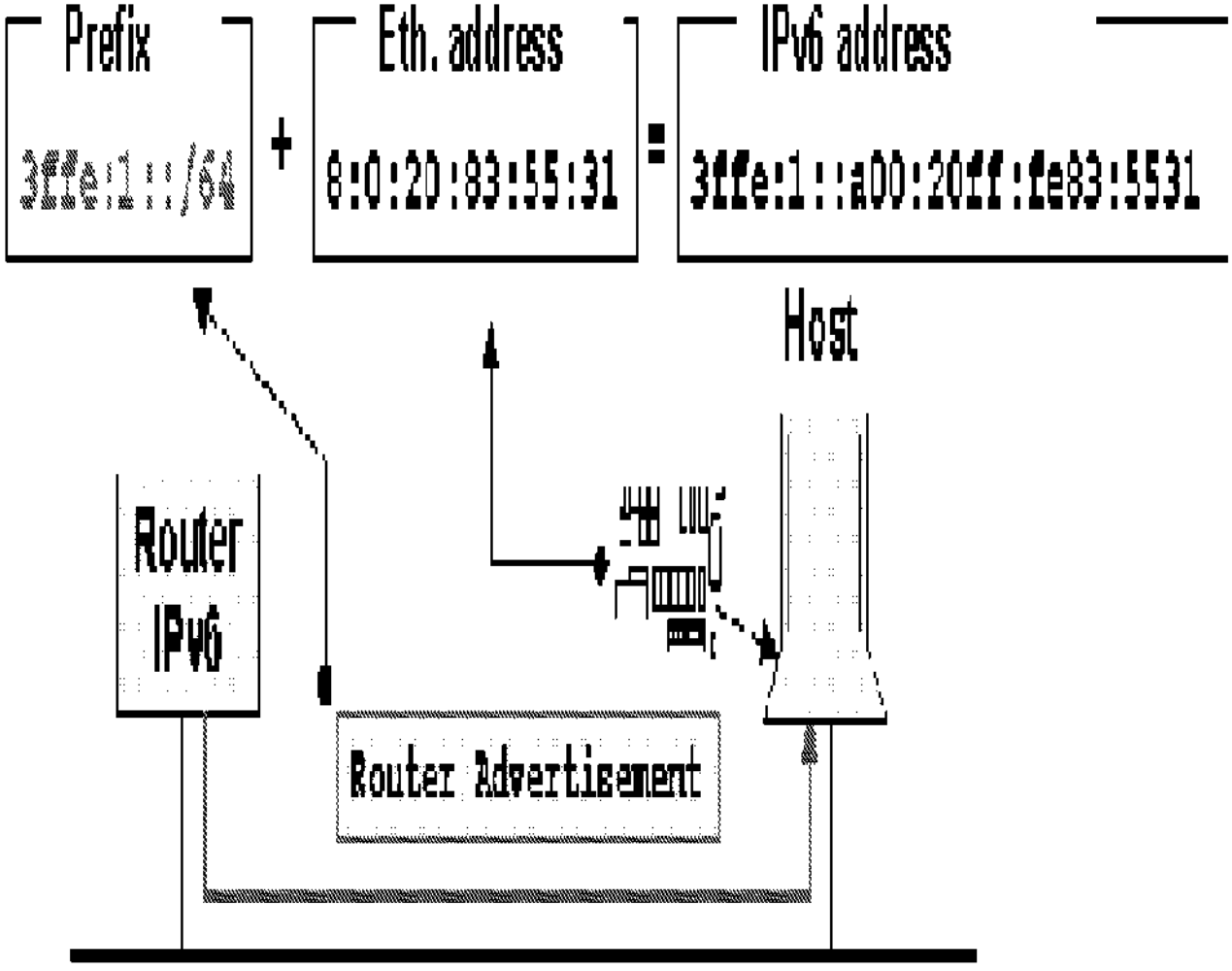
Host autoconfiguration: Host autoconfiguration is a mechanism whereby addresses and other parameters can be assigned to network interfaces. This can be done in two different ways, known as stateful and stateless autoconfiguration. Duplicate Address Detection (DAD) is also performed here.

Router autoconfiguration: Neighbor Discovery protocol the mechanisms for automatic router configuration Keeping a router updated means ensuring that it has an exact knowledge of the organization of the subnet to which it is connected, which in turn means assigning the correct prefixes to each link with which the router has an interface.

DNS autoconfiguration: To facilitate man-machine interfacing, applications generally handle domain names rather than numerical addresses. DNS, database contains name-address mappings for each Internet domain. A6 record type has been defined facilitate the adoption of an automatic DNS management mechanism.

Service autoconfiguration : to make use of the services available on the network, users must know at least the name of the network host on which they are installed. Service Location Protocol (SLP), which provides a flexible and scalable structure whereby hosts can access information concerning the existence, location and configuration of network services.

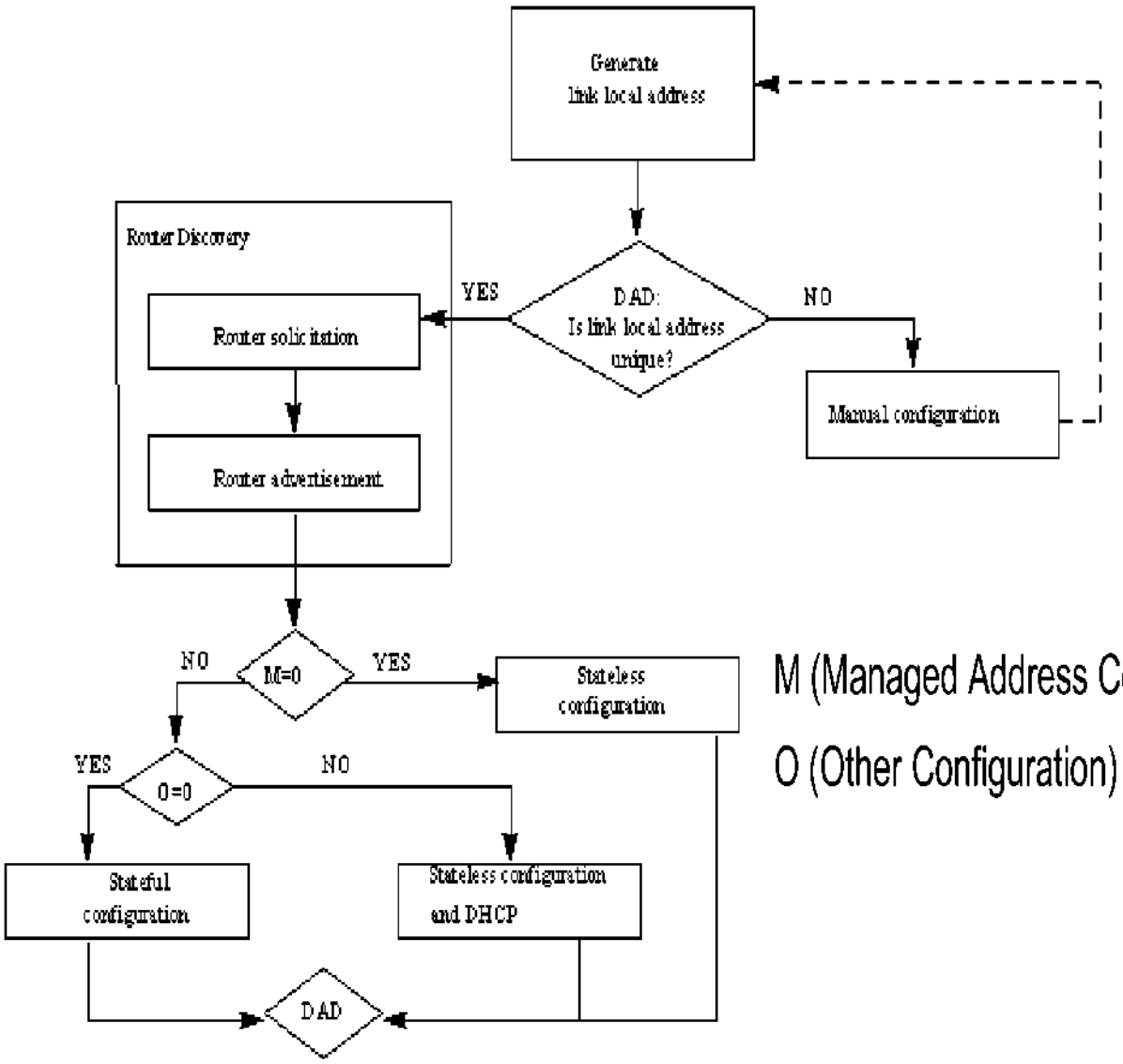
Configuring Network Prefix



WPI



Autoconfiguration Algorithm



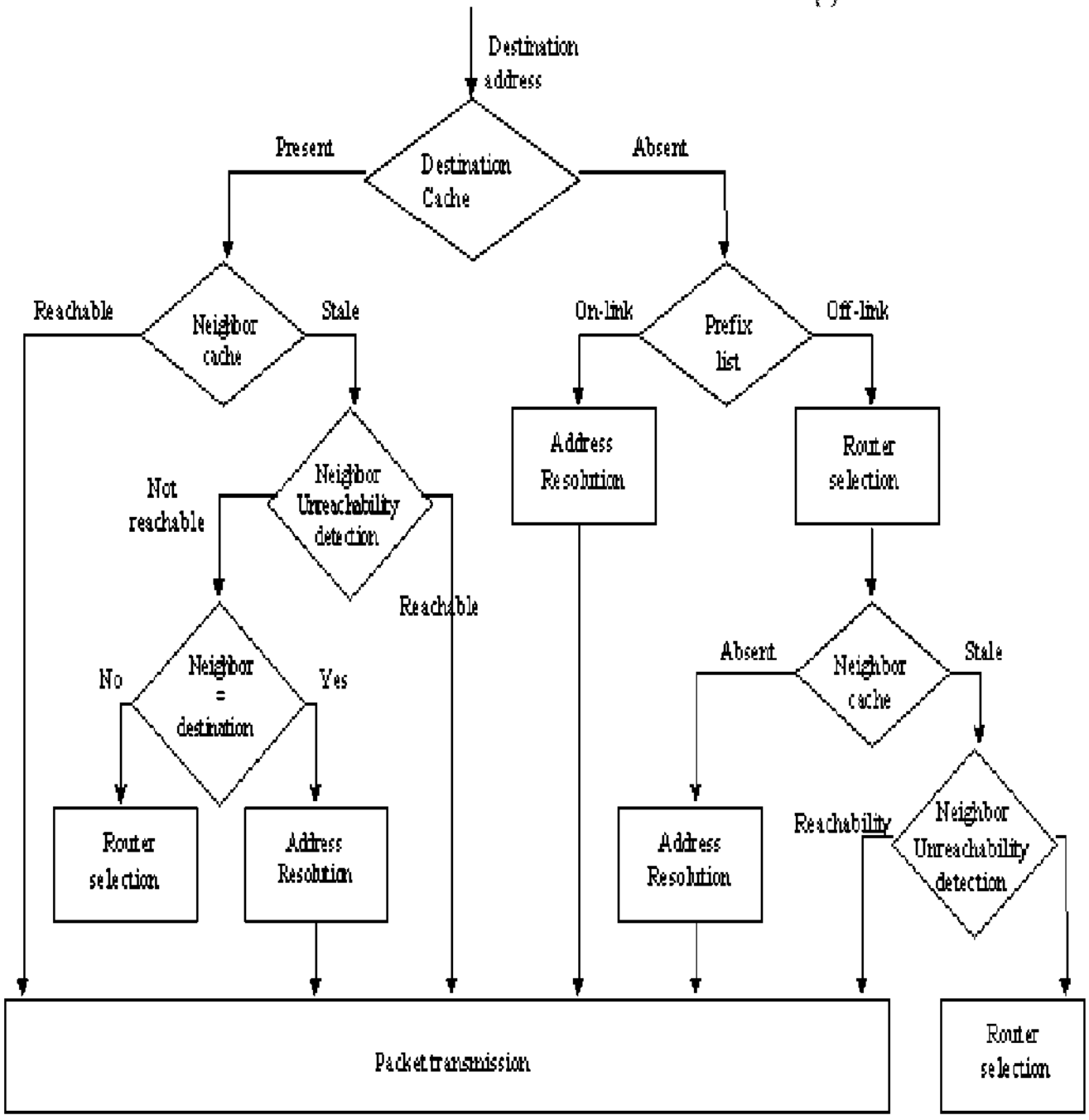
M (Managed Address Configuration)

O (Other Configuration)

WPI



Packet Transmission Algorithm



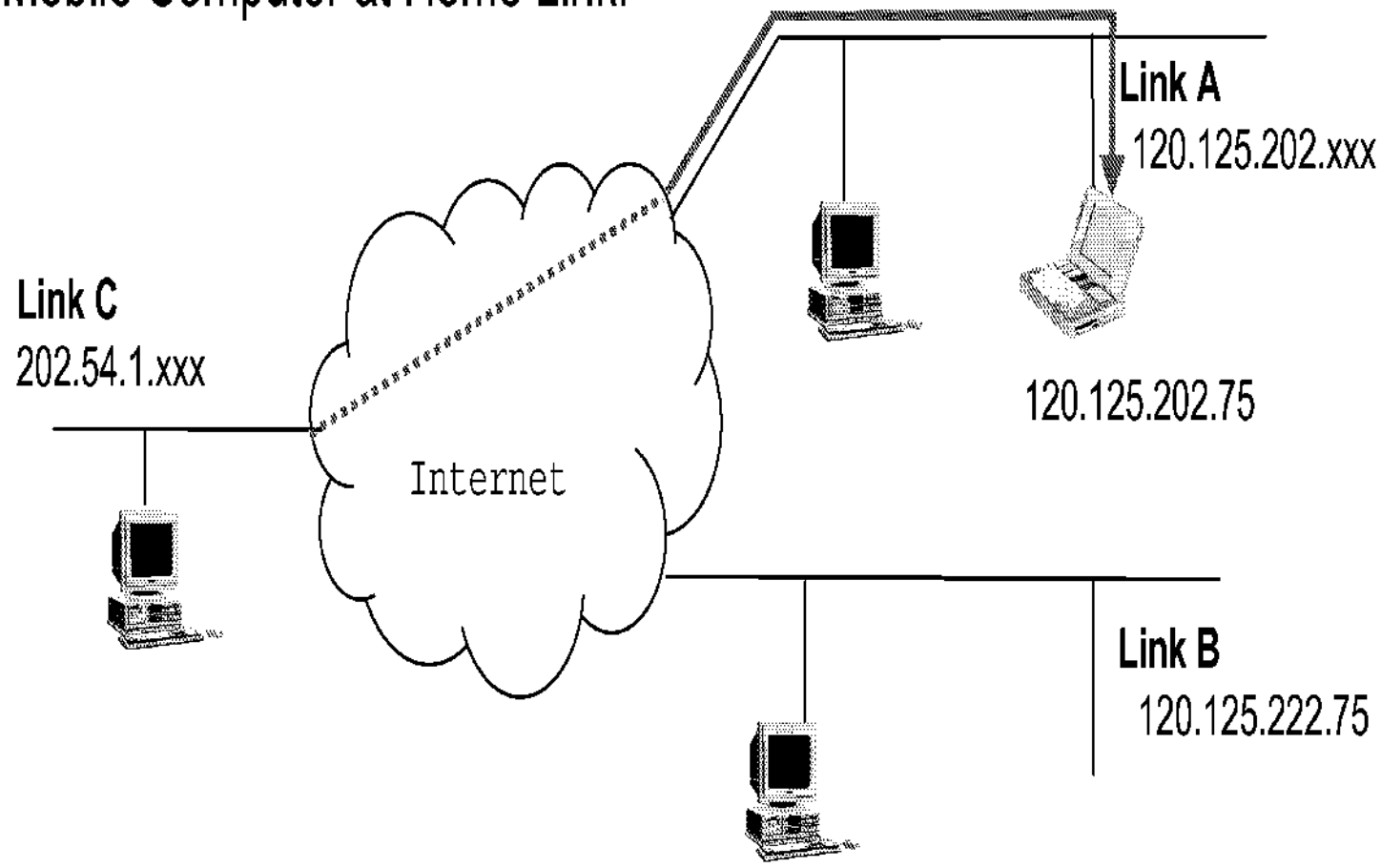
WPI



RED HAT 1010
RED HAT V. CAS

Mobility Problem with IPv4

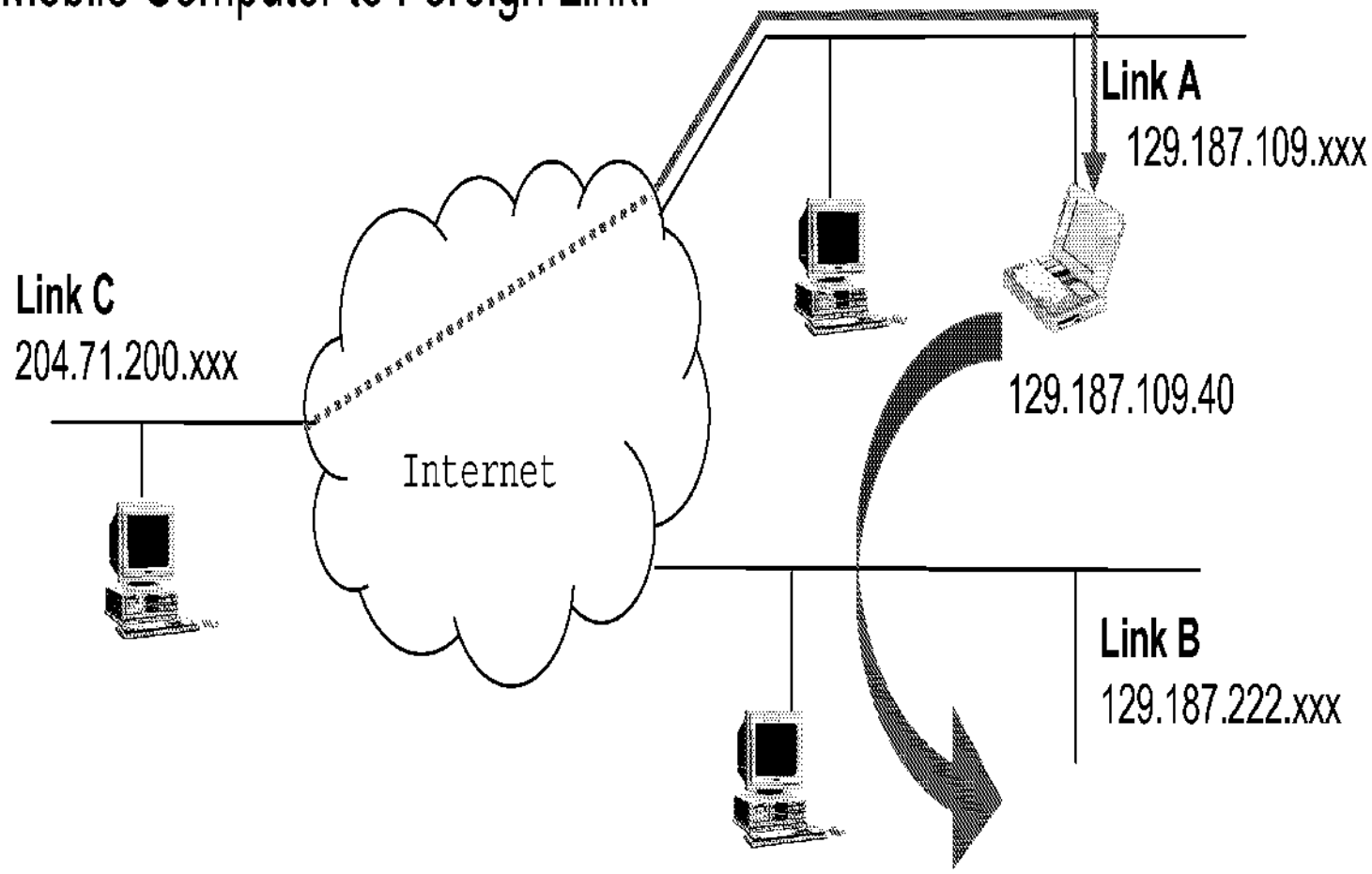
Mobile Computer at Home Link:



WPI

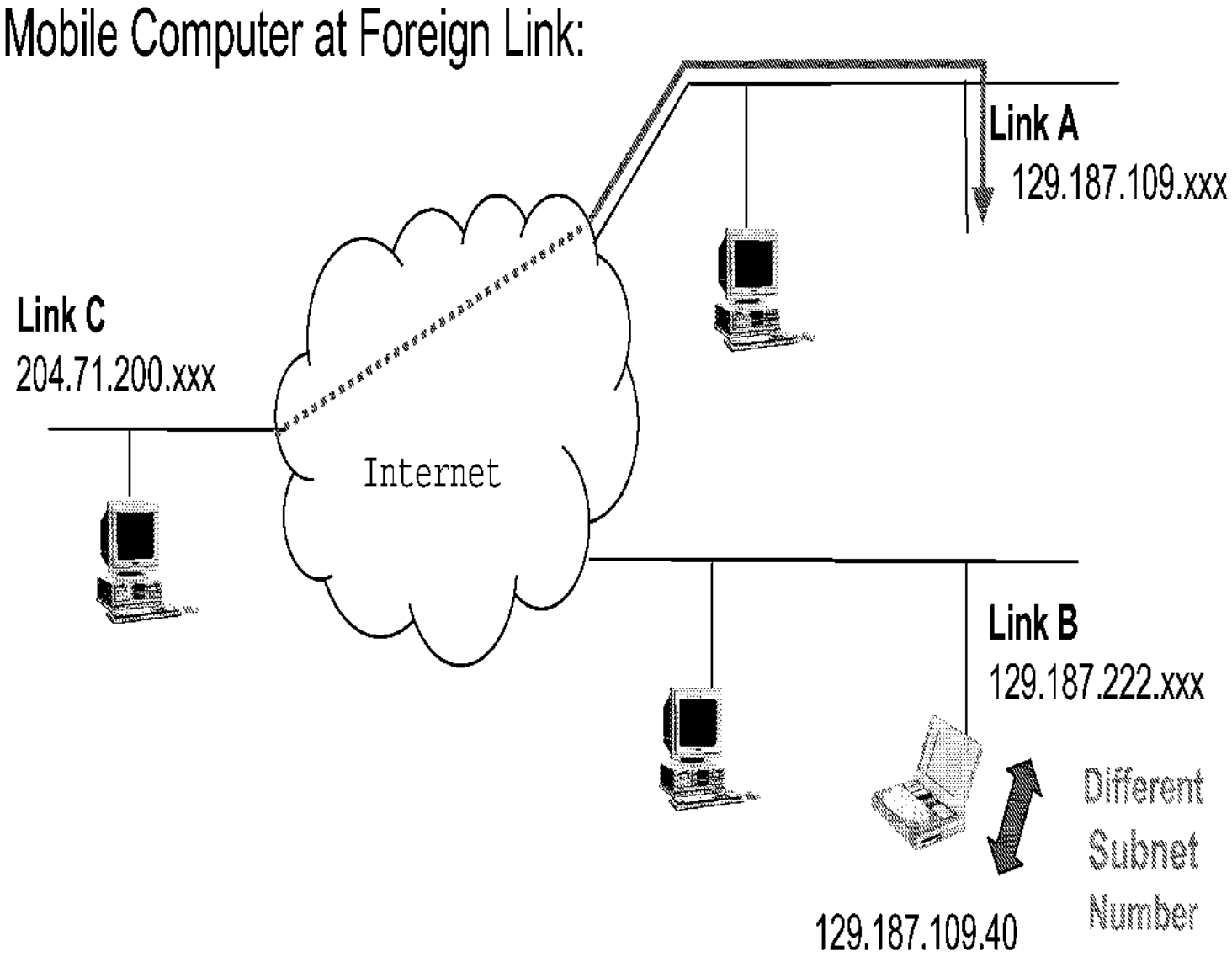
IP Mobility Problem with IPv4

Mobile Computer to Foreign Link:





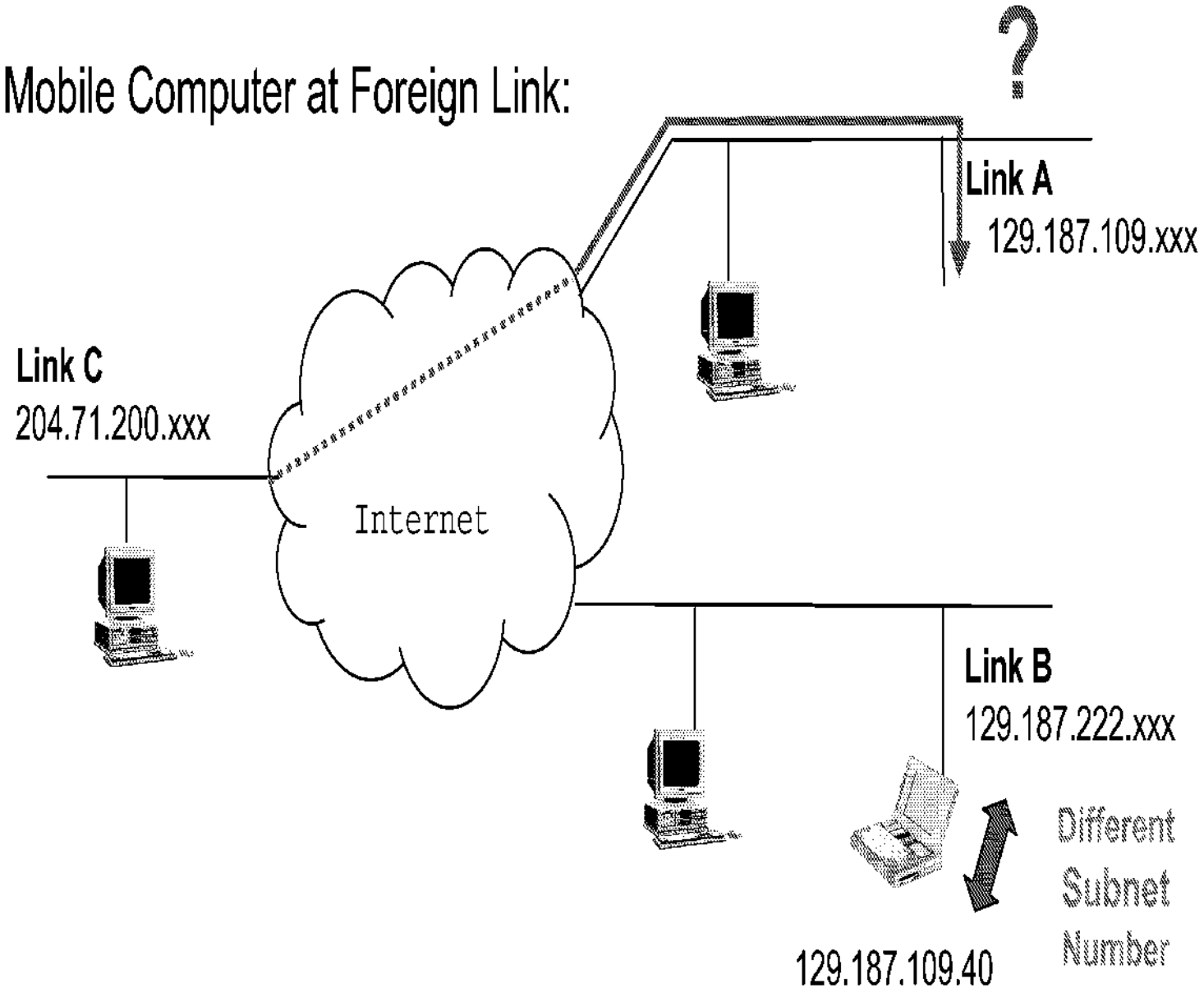
IP Mobility Problem on Movement



WPI

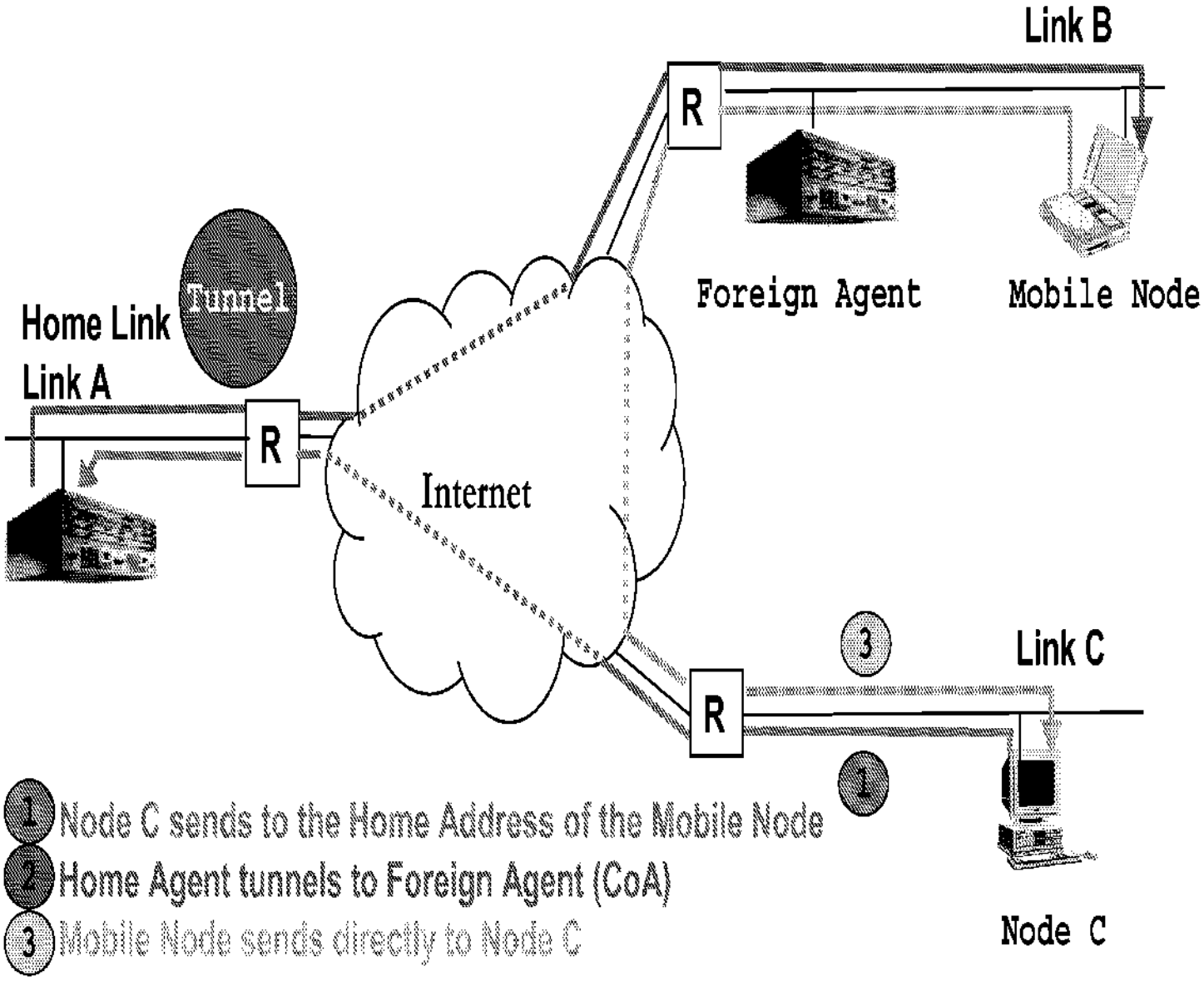
RED HAT 1010
RED HAT V. CAS

IP Mobility Problem with IPv4



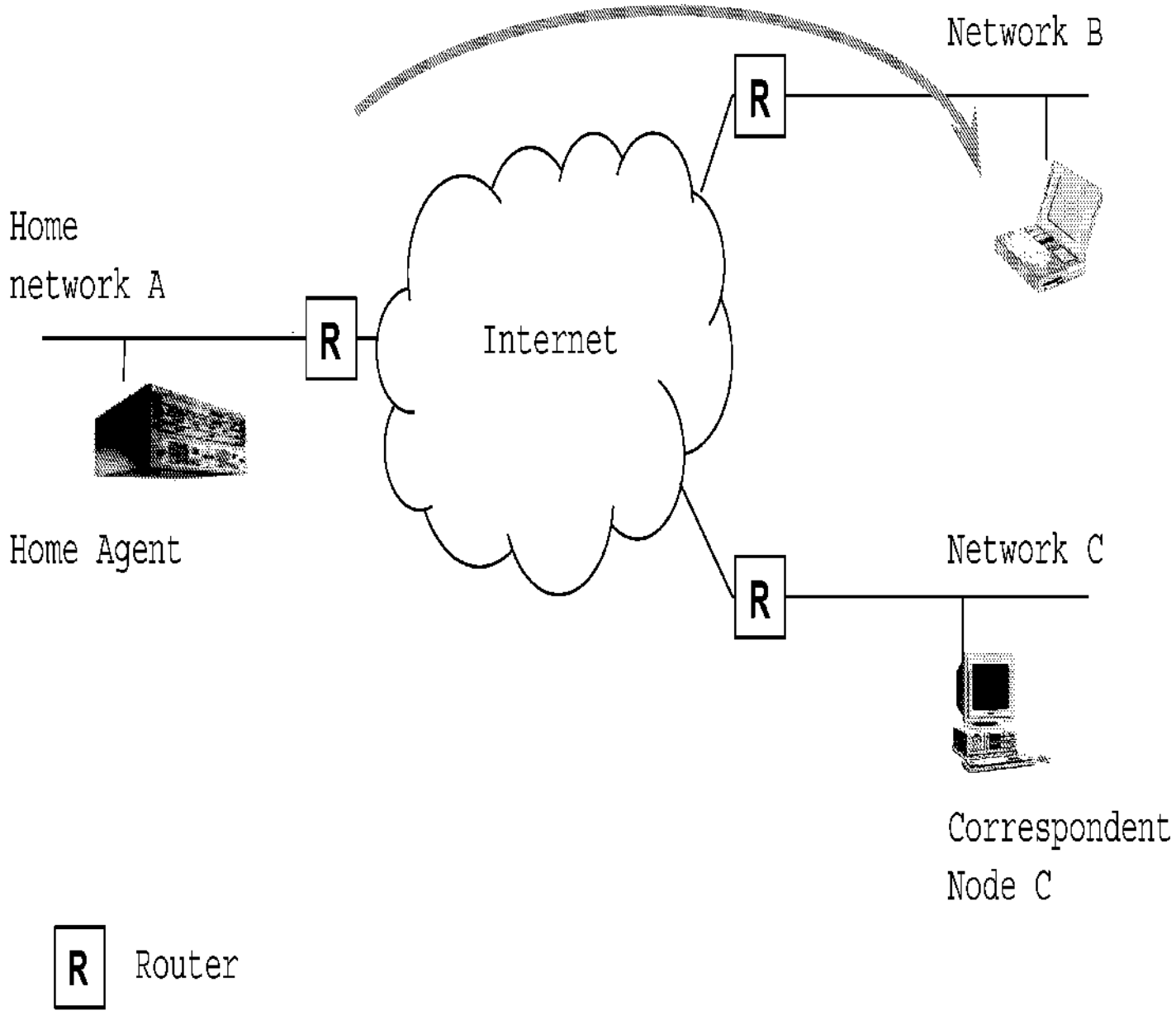
WPI

Packet Delivery with IPv4



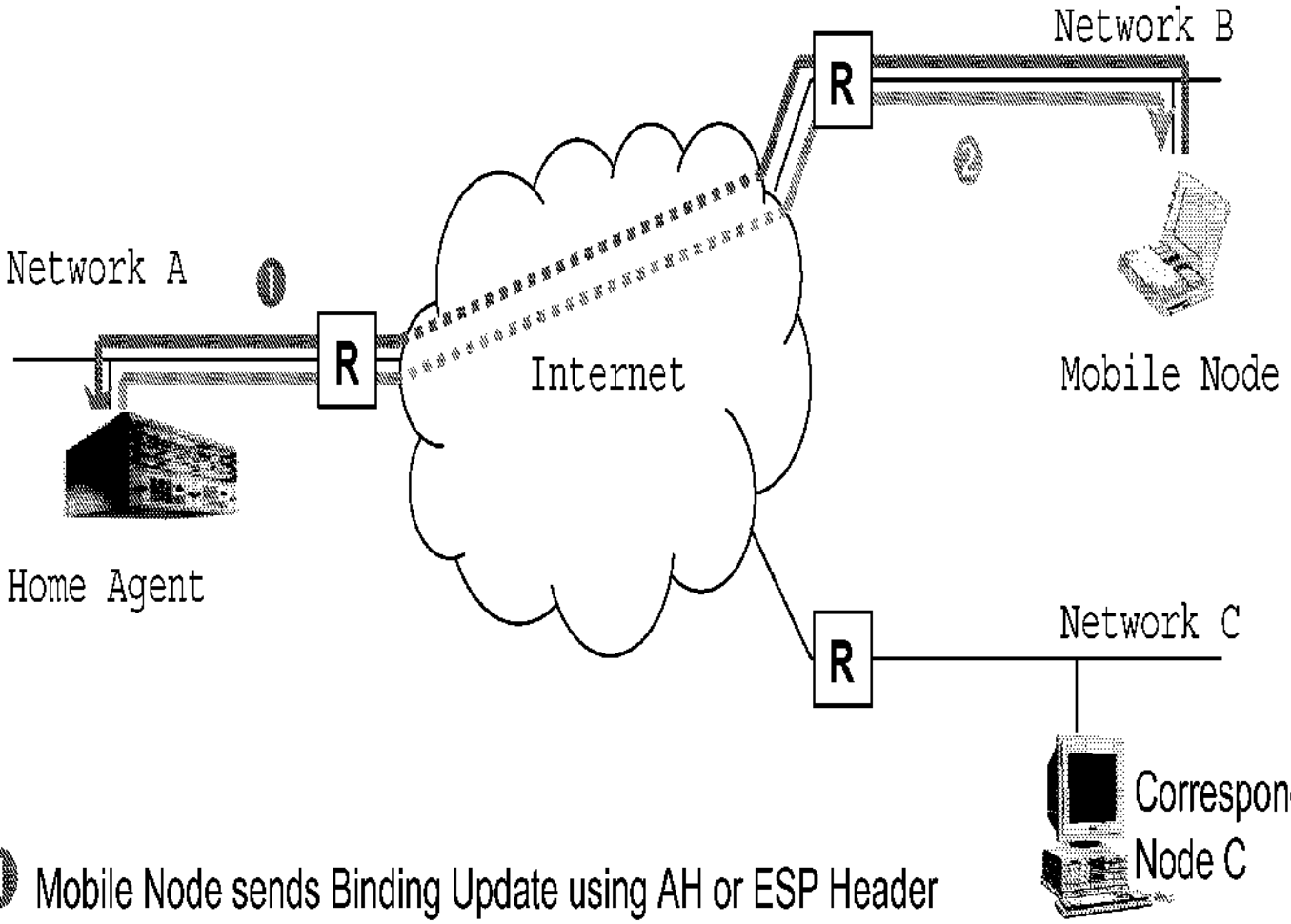
RED HAT 1010
RED HAT V. CAS

Mobile Node Moves: IPv6 consideration



RED HAT 1010
RED HAT V. CAS

Mobile Node registers at its Home Agent



1. Mobile Node sends Binding Update using AH or ESP Header
2. Home Agent replies with Binding Acknowledgement using AH or ESP Header

WPI



RED HAT 1010
RED HAT V. CAS

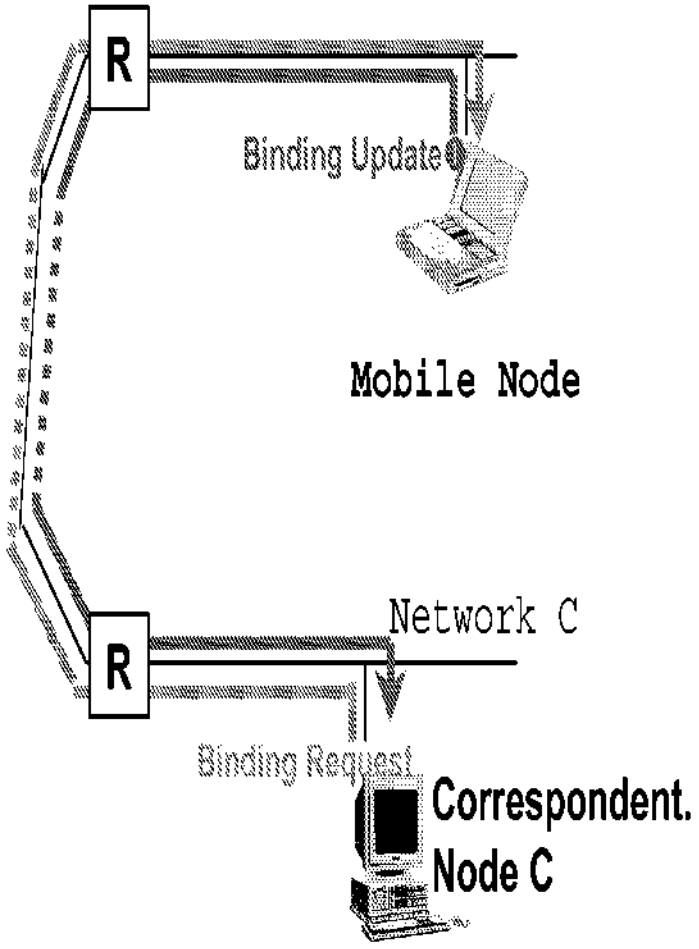
Binding Request

When Mobile's Node Care-of address lease-time going to expire.

To keep Correspondent Node update.

Mobile Node Send Binding Update

Correspondent Node Request Binding Update



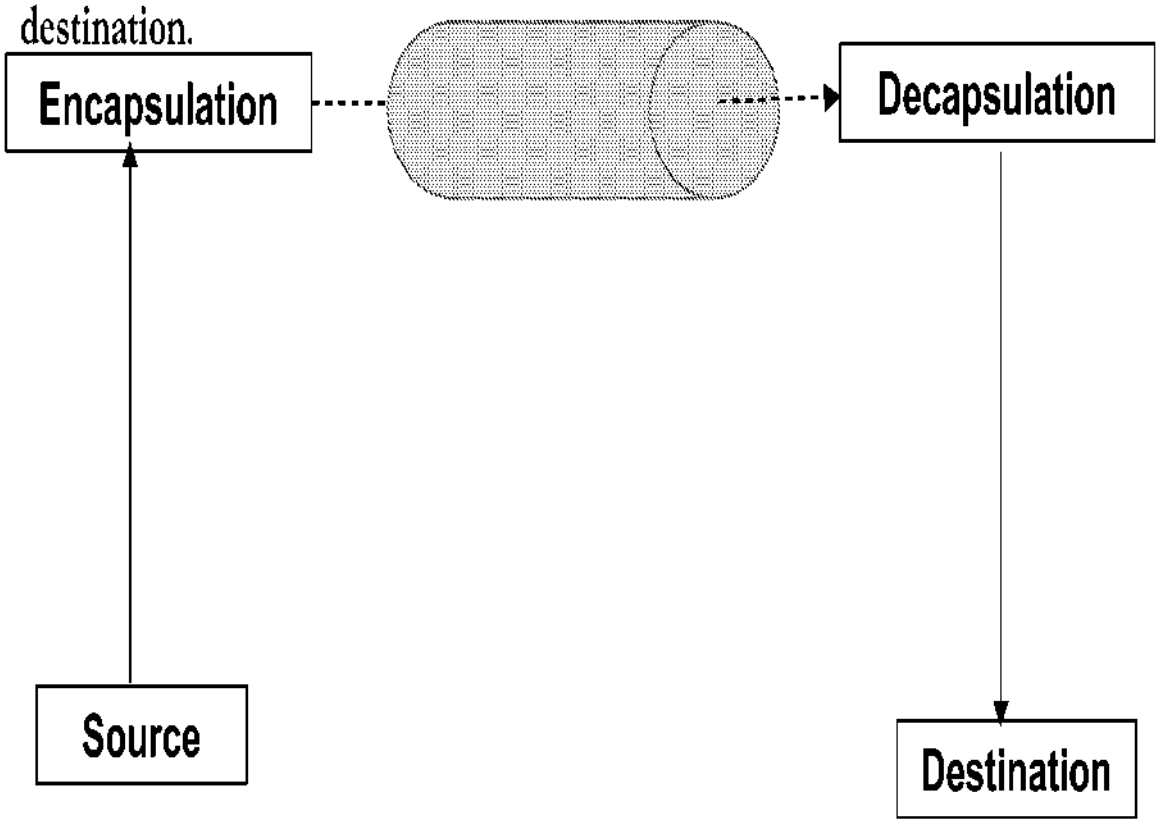
WPI



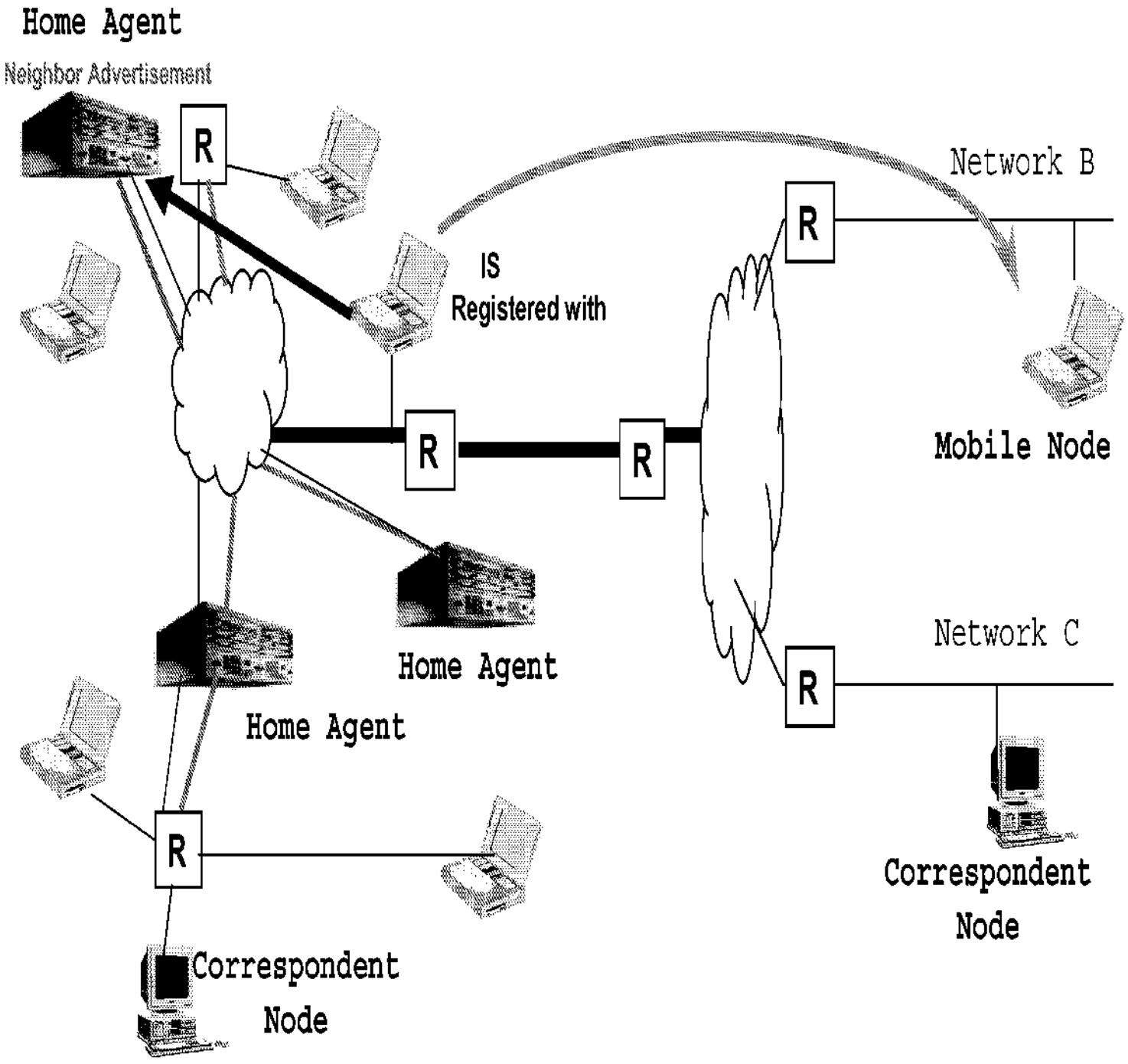
RED HAT 1010
RED HAT V. CAS

Tunneling

Tunnel: The path followed by a datagram while it is encapsulated. While encapsulated, a datagram is routed to a knowledgeable agent, which decapsulates the datagram and then forwards it to its ultimate destination.



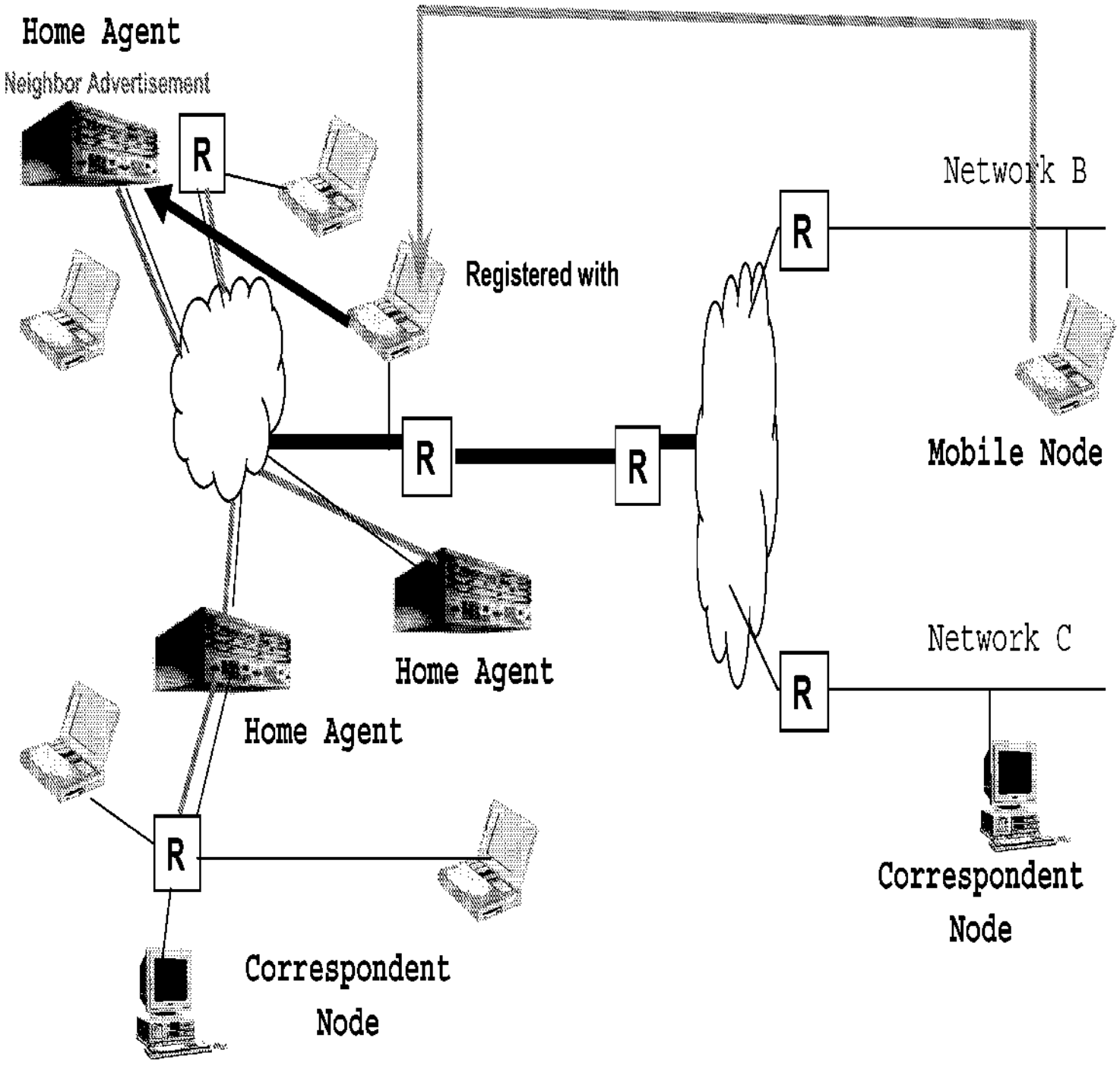
On Mobile Node Movement: HA Takes Action



WPI

RED HAT 1010
RED HAT V. CAS

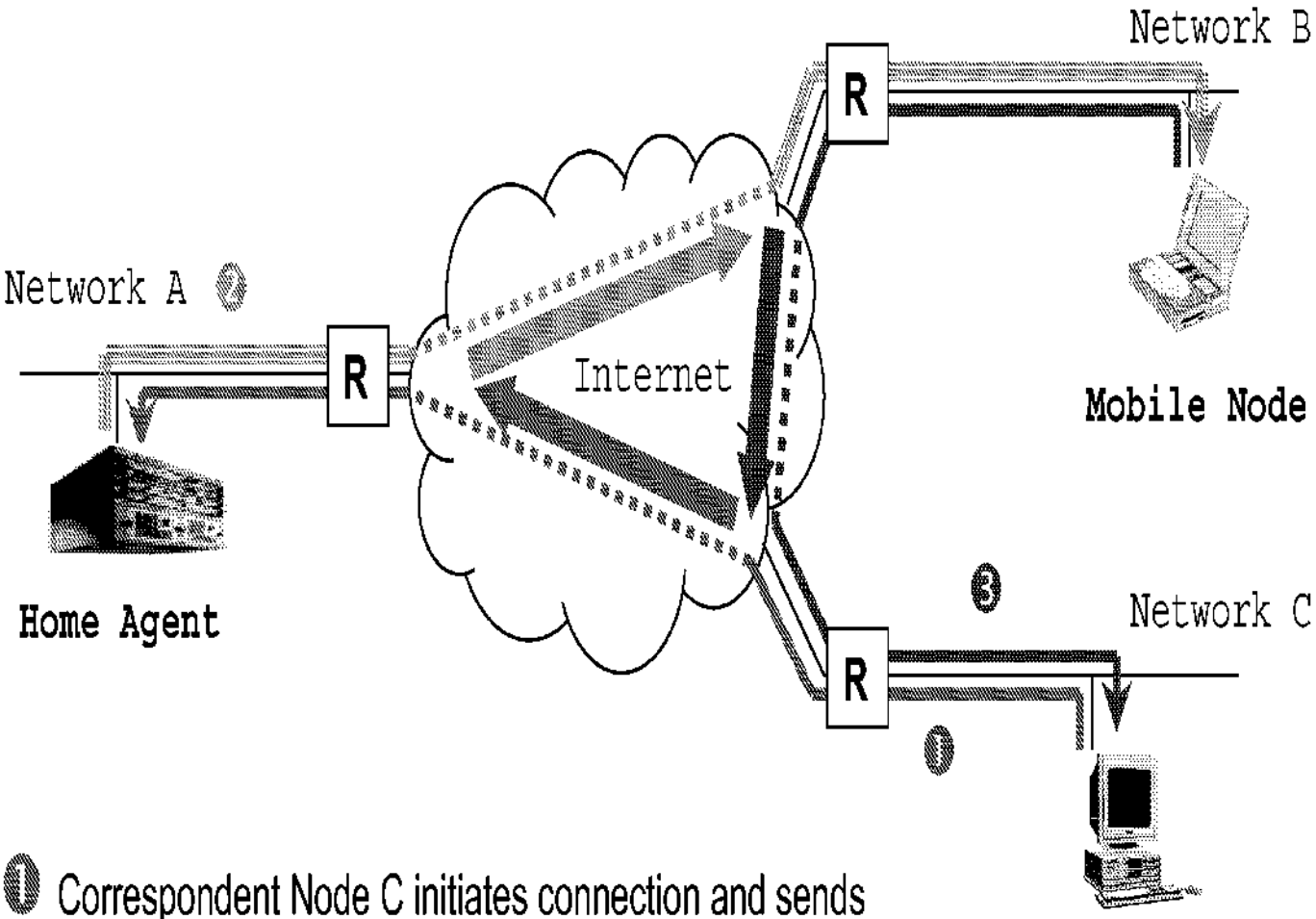
HA Takes Action: When MN Return its Home Subnet



WPI

RED HAT 1010
RED HAT V. CAS

Triangular Routing during Initial Phase



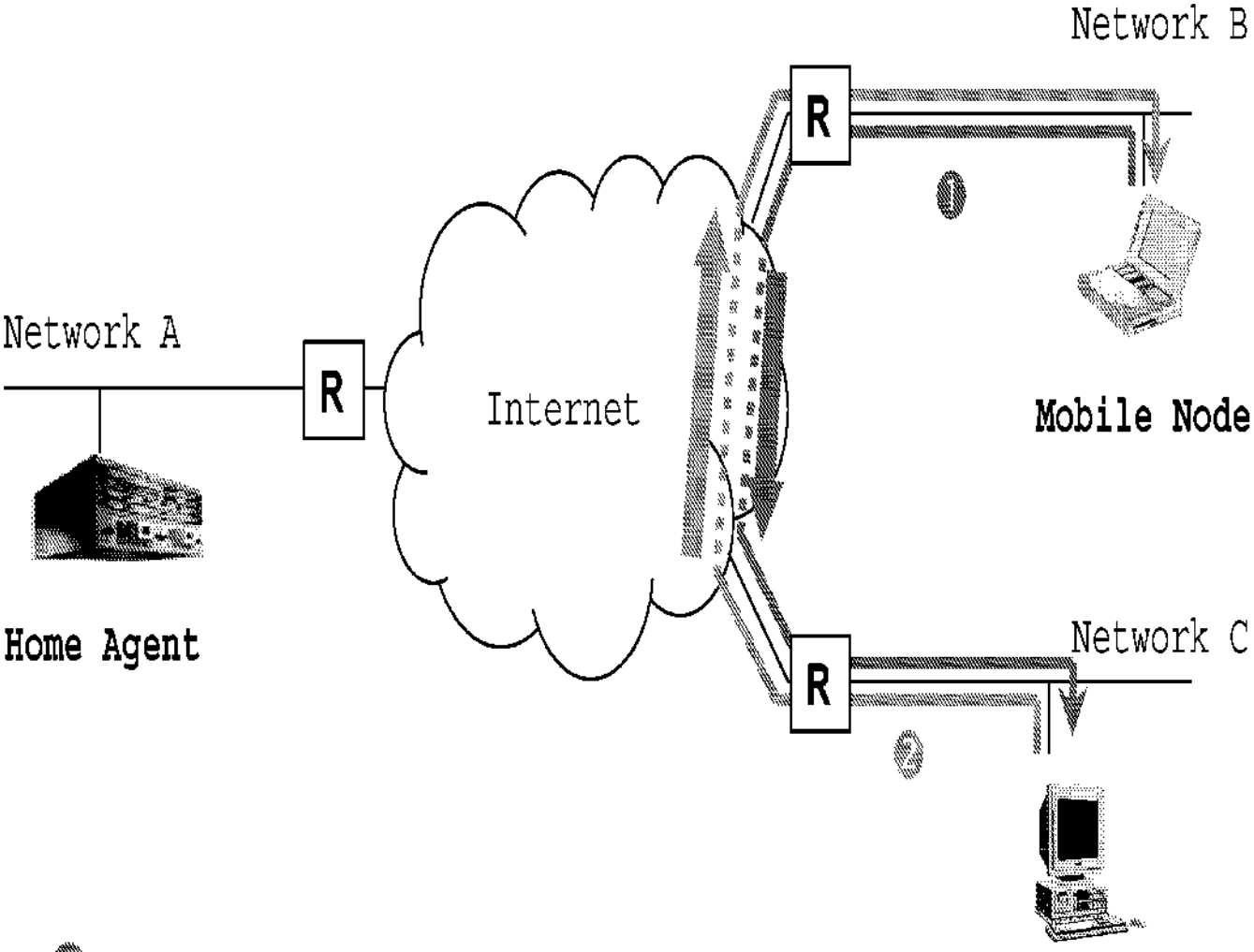
1. Correspondent Node C initiates connection and sends packets to the Home Address of the Mobile Node
2. Home Agent intercepts packets and tunnels them to the Mobile Node
3. Mobile Node sends answer directly to Host C

WPI



RED HAT 1010
RED HAT V. CAS

Normal Operation by Route Optimization



Correspondent Node

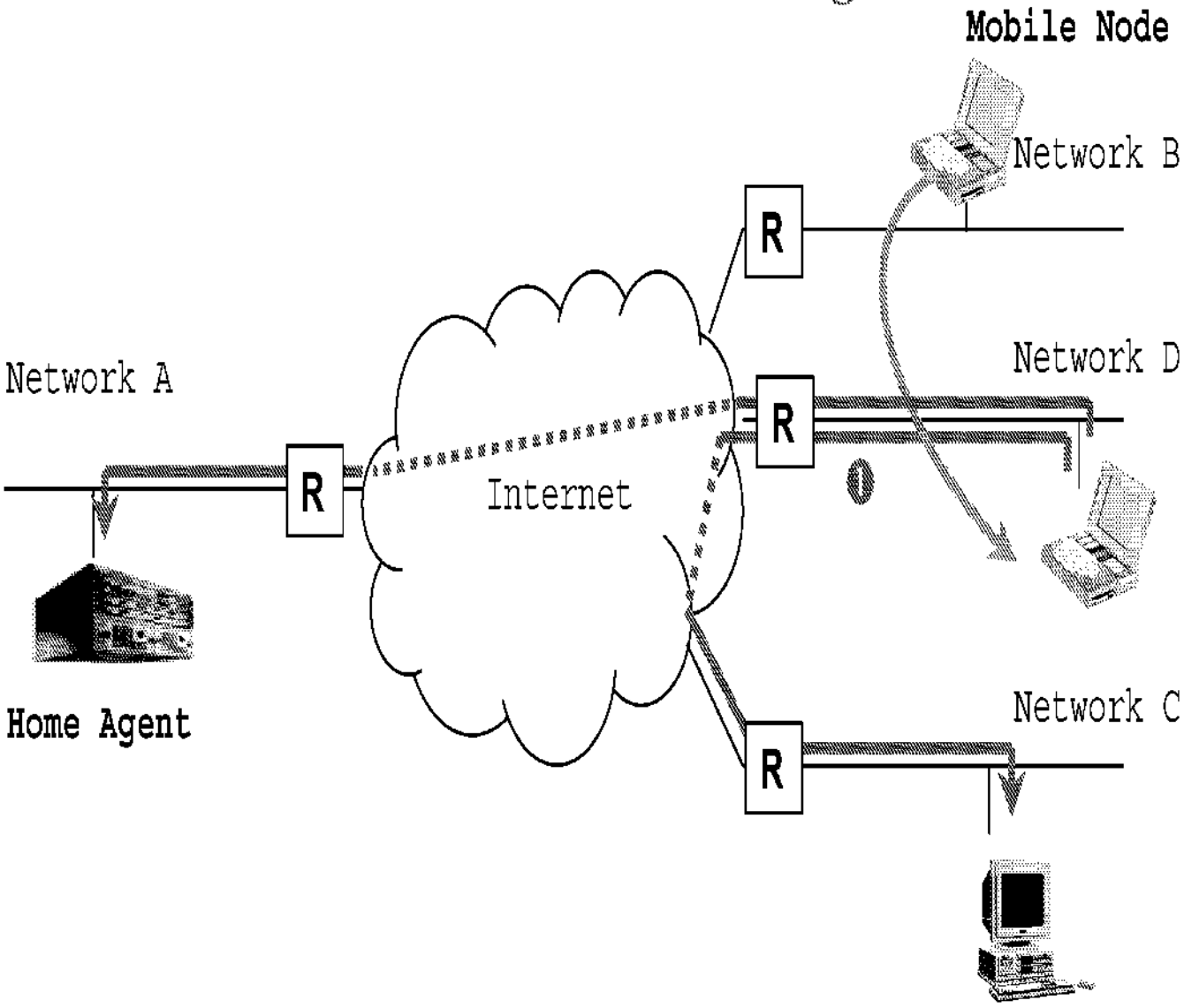
1. Mobile Node sends Binding Update to Correspondent Node C
2. Now Correspondent Node can address the CoA of the Mobile Node directly

WPI



RED HAT 1010
RED HAT V. CAS

Mobile IPv6 Roaming



1. Mobile Node sends Binding Updates to the Home Agent and to all the Nodes, he is connected to

WPI



RED HAT 1010
RED HAT V. CAS

Movement Detection

Scenario-I : Mobile node to know quickly when the when the Default router will be unavailable



Neighbor Advertisement unreachable detection by using upper-layer TCP time-out mechanism.

When Mobile node don't receive Neighbor Advertisement Message from default router in response to Neighbor Solicitation message.

Scenario II : When Mobile node become unreachable to default Router



Some sort of time setting its network interface so that it can receive all the packets through that router.

Through Router Advertisement messages. receipt of packets from default router indicate reachable.



RED HAT 1010
RED HAT V. CAS

Renumbering Home Subnet

When Home subnet change its internet service from different ISP then its Network Prefix changes (thereby Network Prefix of all nodes on Home Subnet also changes)

Nodes on the Home Subnet update their Network Prefix, via Neighbor Discovery mechanism.

Node which is away from Home Subnet need Special Care.

Home Agent tunnel Authenticated Router Advertisement to each Mobile node it serve.

Mobile node performs standard autoconfiguration mechanism to create new Home Address.

When Mobile node return home, it first performs duplicate address detection.

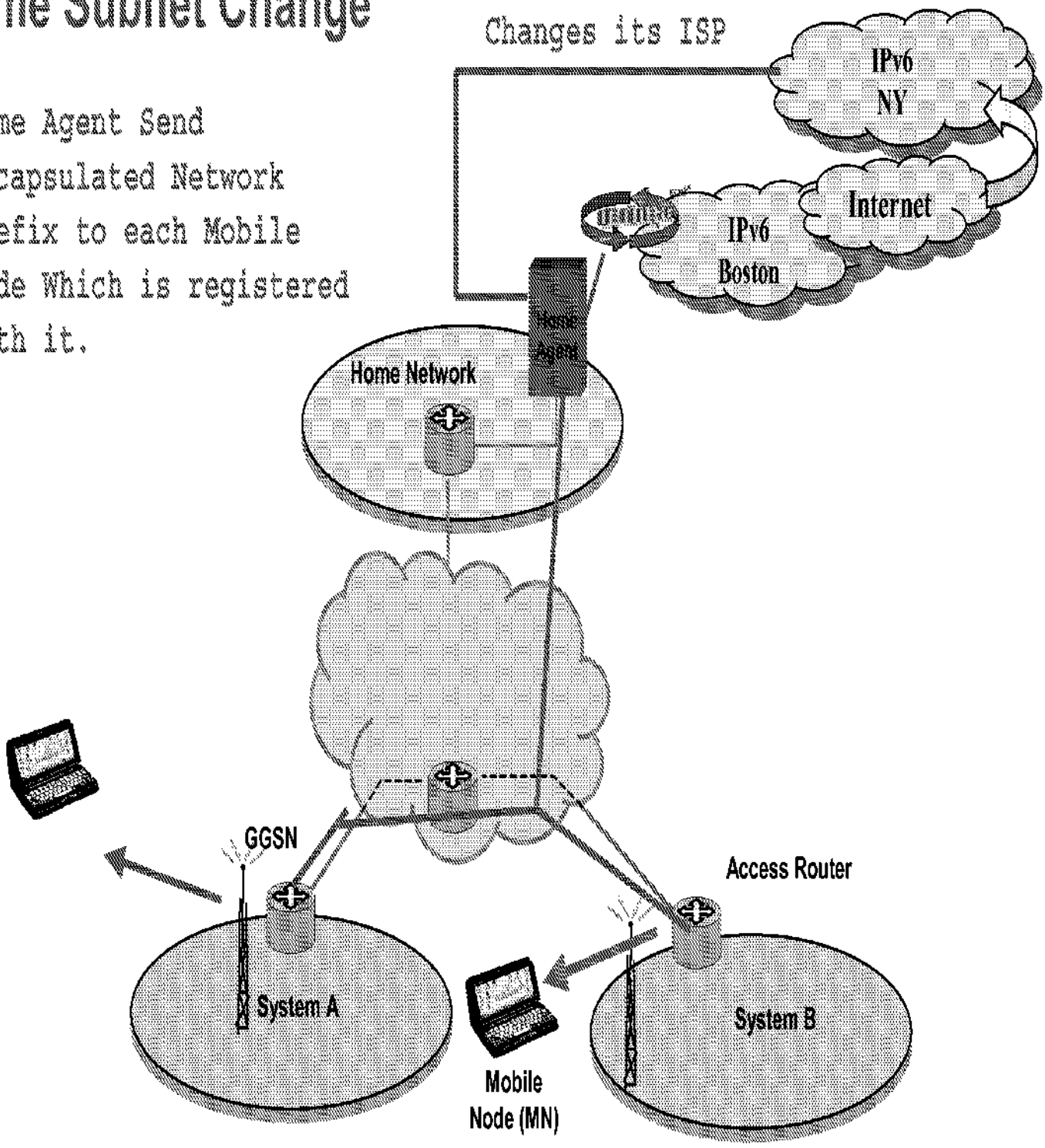
WPI



RED HAT 1010
RED HAT V. CAS

Home Subnet Change

Home Agent Send Encapsulated Network Prefix to each Mobile Node Which is registered with it.



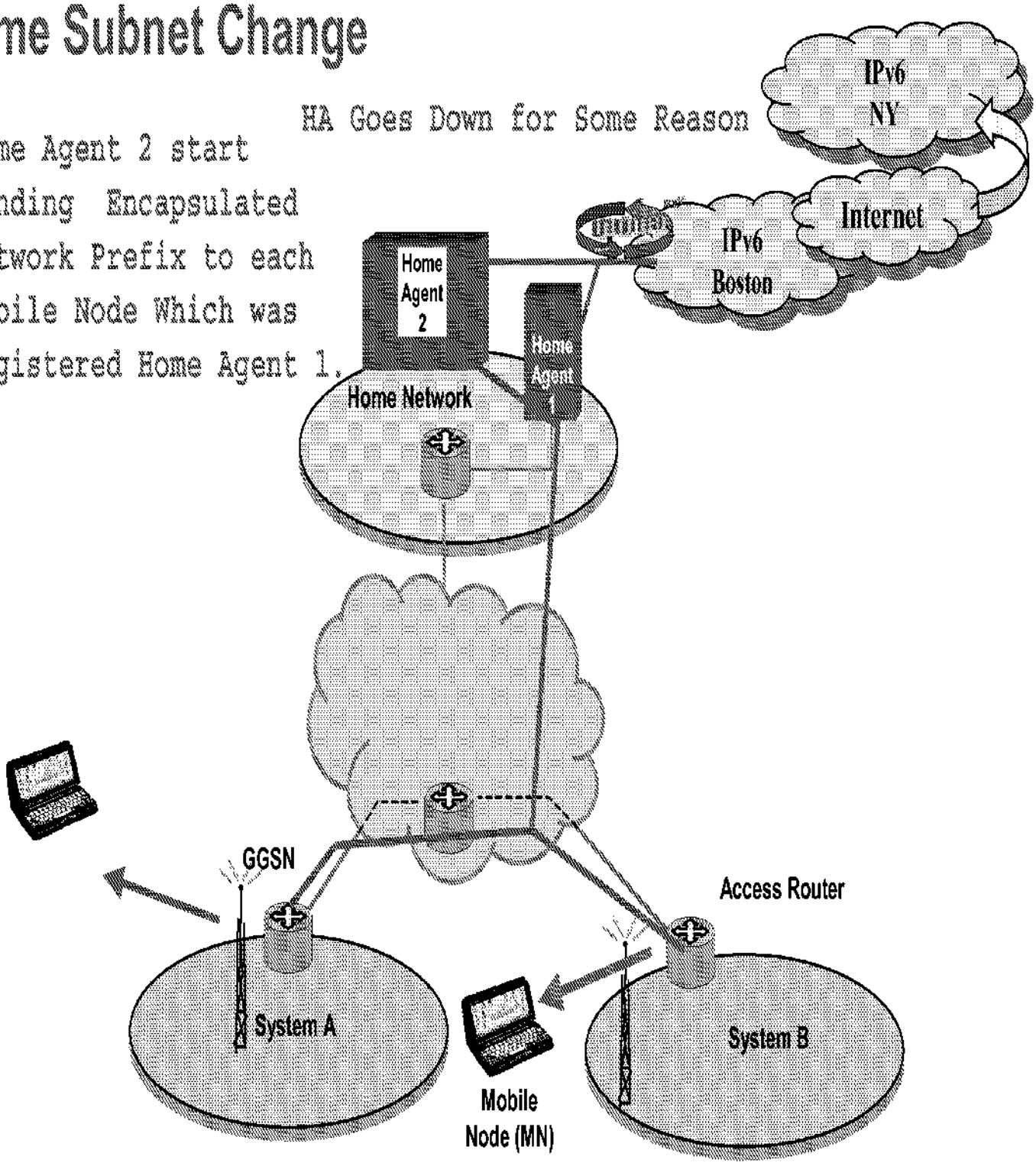
WPI

RED HAT 1010
RED HAT V. CAS

Home Subnet Change

HA Goes Down for Some Reason

Home Agent 2 start Sending Encapsulated Network Prefix to each Mobile Node Which was registered Home Agent 1.

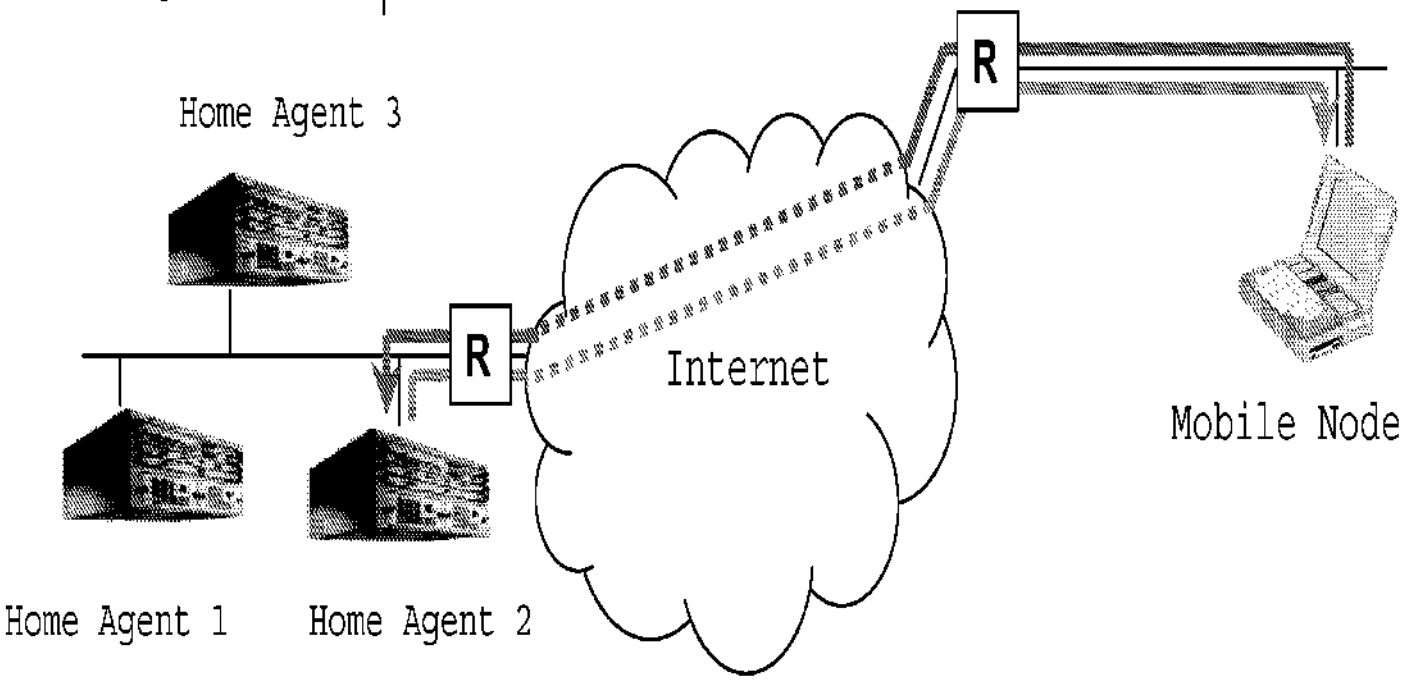


WPI

RED HAT 1010
RED HAT V. CAS

Dynamic Home Agent Address Discovery (Renumbering)

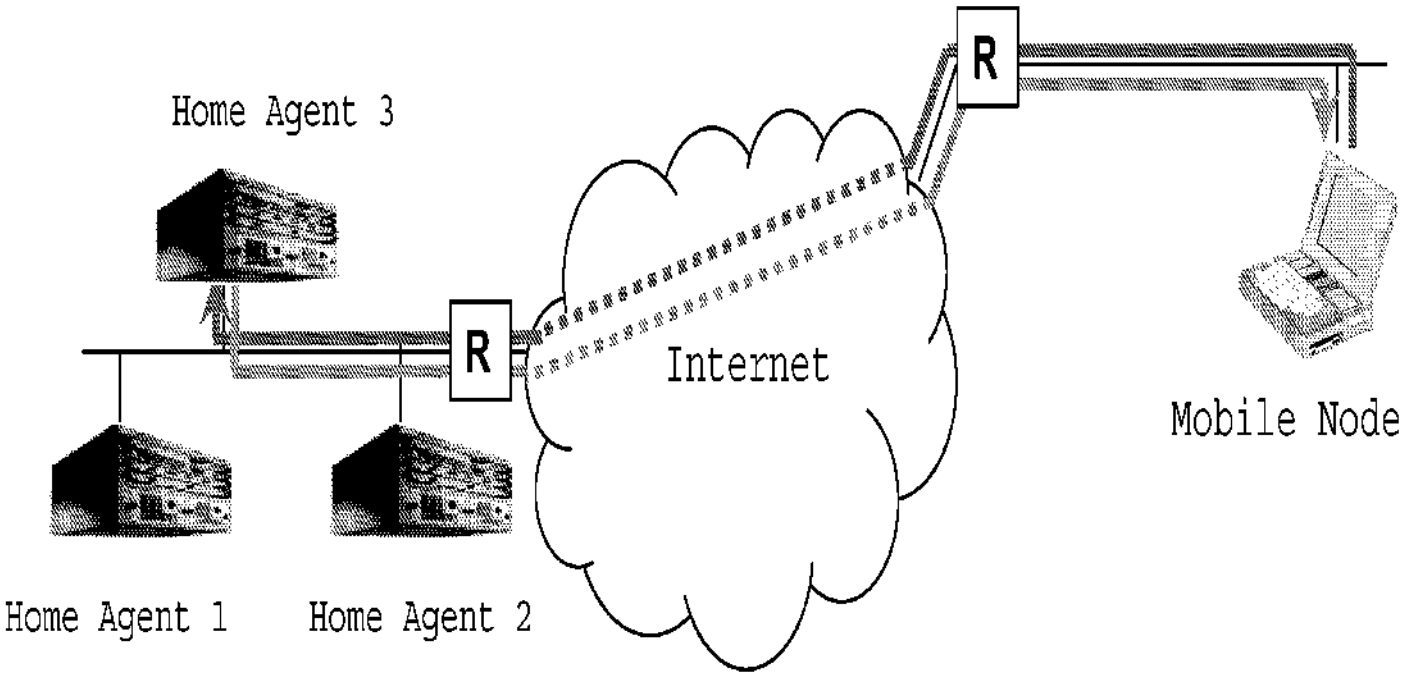
Home Agents List	Priority
Home Agent 3	9
Home Agent 1	2
Home Agent 2	-3



1. Mobile Node sends Binding Update to the Home Agents Anycast Address of its home network
2. One Home Agent answers with Binding Acknowledgement containing a list of available Home Agents

Registration at selected Home Agent

Home Agents List	Priority
Home Agent 3	9
Home Agent 1	2
Home Agent 2	-3



1. Mobile Node sends Binding Update to the first Home Agent contained in the Home Agents List
2. Binding Acknowledgement completes Registration process

WPI

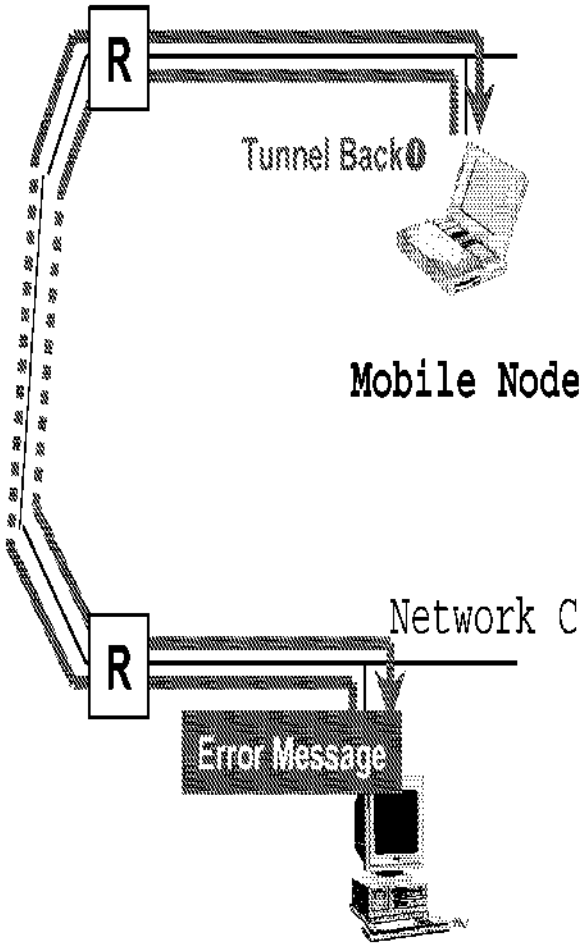


RED HAT 1010
RED HAT V. CAS

ICMP Role

When an IPv6 node discards a packet, it sends an error message to the source. There are four types of message:

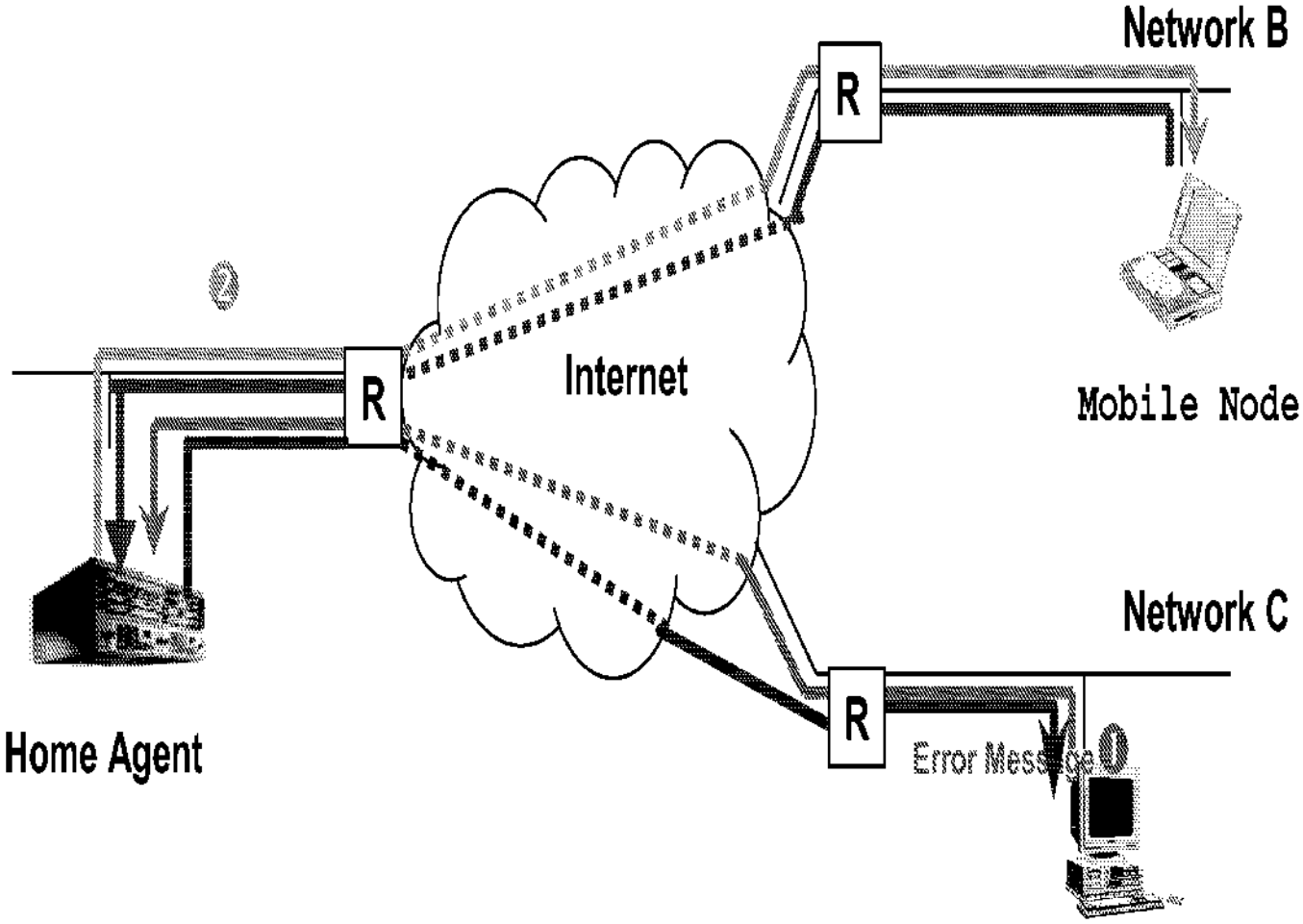
1. Destination unreachable (type=1). Sent by a router to the source when a packet cannot be forwarded to its destination.
2. Packet too big (type =2). Used when the link MUT on the forwarding link is smaller than the packet.
3. Time exceeded (type=3). Indicates that the packet's hop limit field is zero.
4. Parameter problem (type=4). Indicates that a field of the datagram is not recognized as valid and the packet can thus not be processed.



ICMP: Includes the so-called Neighbor Discovery mechanisms, the terminal autoconfiguration mechanisms and address resolution mechanisms.

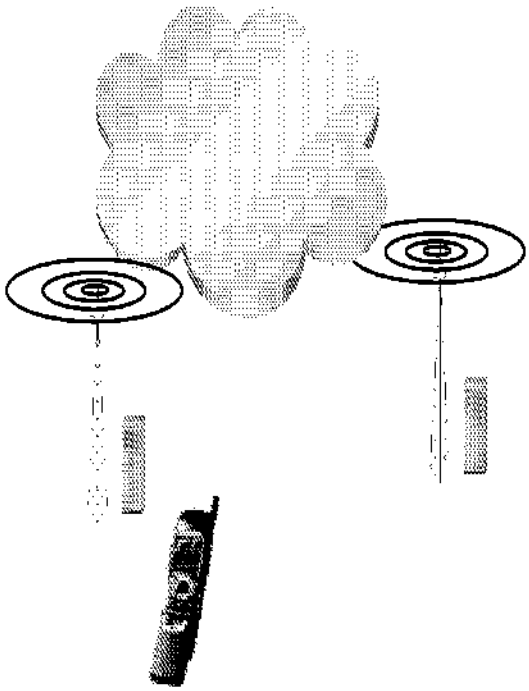
Handling ICMP Scenario 2

When CoN send error message through Home Agent.

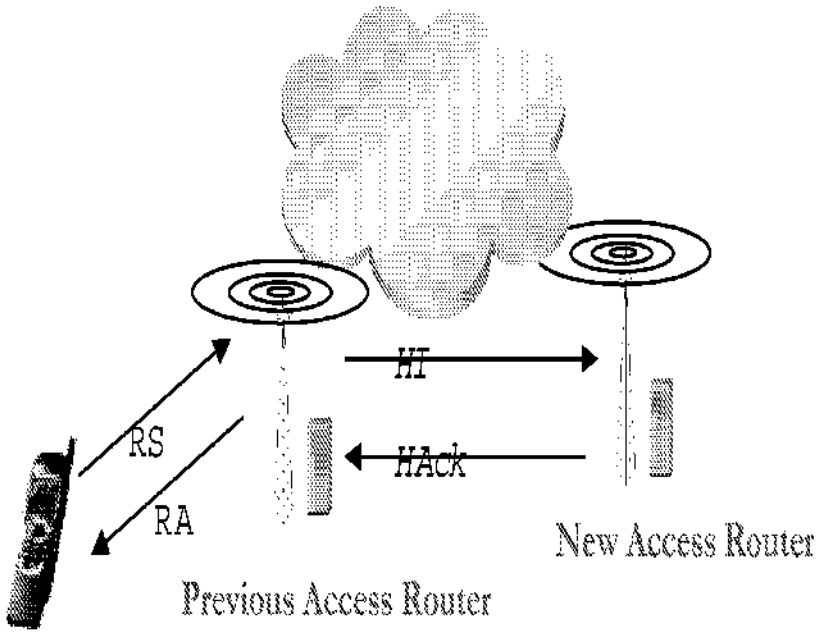


Smooth/Fast/Seamless Handover

- Smooth handover == low loss
- Fast handover == low delay
  - 30 ms?
  - Duplicate Address Detection?? (can router pre-empt this?)
- Seamless handover == *smooth* and *fast*



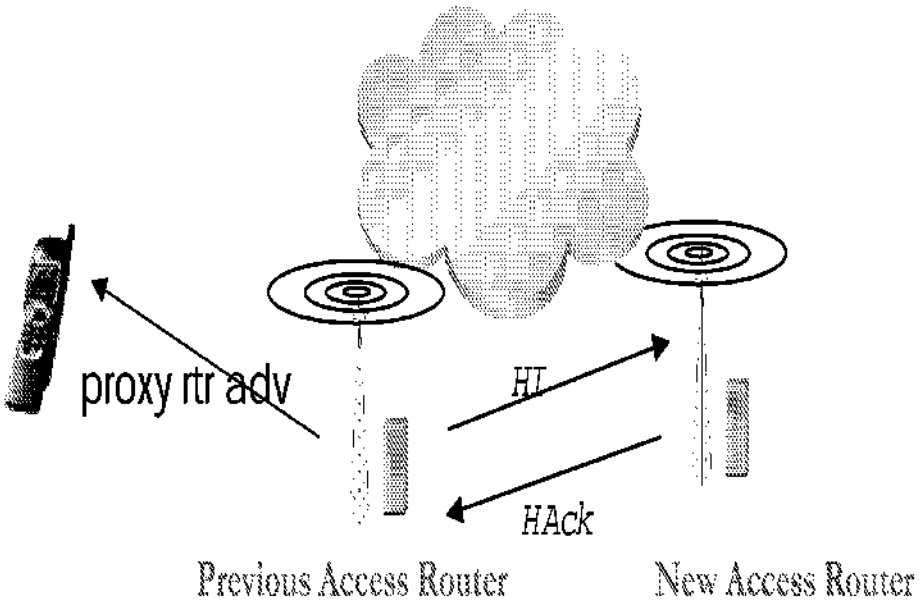
Mobile-controlled seamless handover



One scenario: mobile sends special Router Solicitation (RS)

- Previous Access Router replies with *Proxy* Router Advert. (RA)
- Previous Access Router sends Handover Initiate (HI)
- New Access Router sends Handover Acknowledge (HACK)

Network Controlled Handover



- Previous access router sends Proxy Router Advertisement on behalf of the new access router – contains prefix and lifetime information, etc.
- Previous access router sends *Handover Initiate* message to new access router
- Mobile node *MAY* finalize context transfer at new access router



RED HAT 1010
RED HAT V. CAS

Ongoing Work for Open Questions

Security issues: **Firewalls, cause difficulty for Mobile IP because they block all classes of incoming packets that do not meet specified criteria.**

→ Gupta and Glass have proposed a firewall traversal extend Mobile IP operation across firewalls, even when multiple security domains are involved.

Ingress filtering: **Many border routers discard packets if the packets do not contain a source IP address configured for one of the enterprise's internal networks**

→ Montenegro has proposed the use of reverse tunnels to the home agent to counter the restriction imposed by ingress filtering.

Deficiency **of Mobile IPv6, is that it does not support fast handoff – (this is the ability to switch to another subnet without significant delay or loss of packets). Excessive signalling in rapidly changing cells.**

⇒ Extension to Mobile IPv6 called "**HIERARCHICAL MOBILE IP v6**".



RED HAT 1010
RED HAT V. CAS

Hierarchical Mobile IPv6

Hierarchy

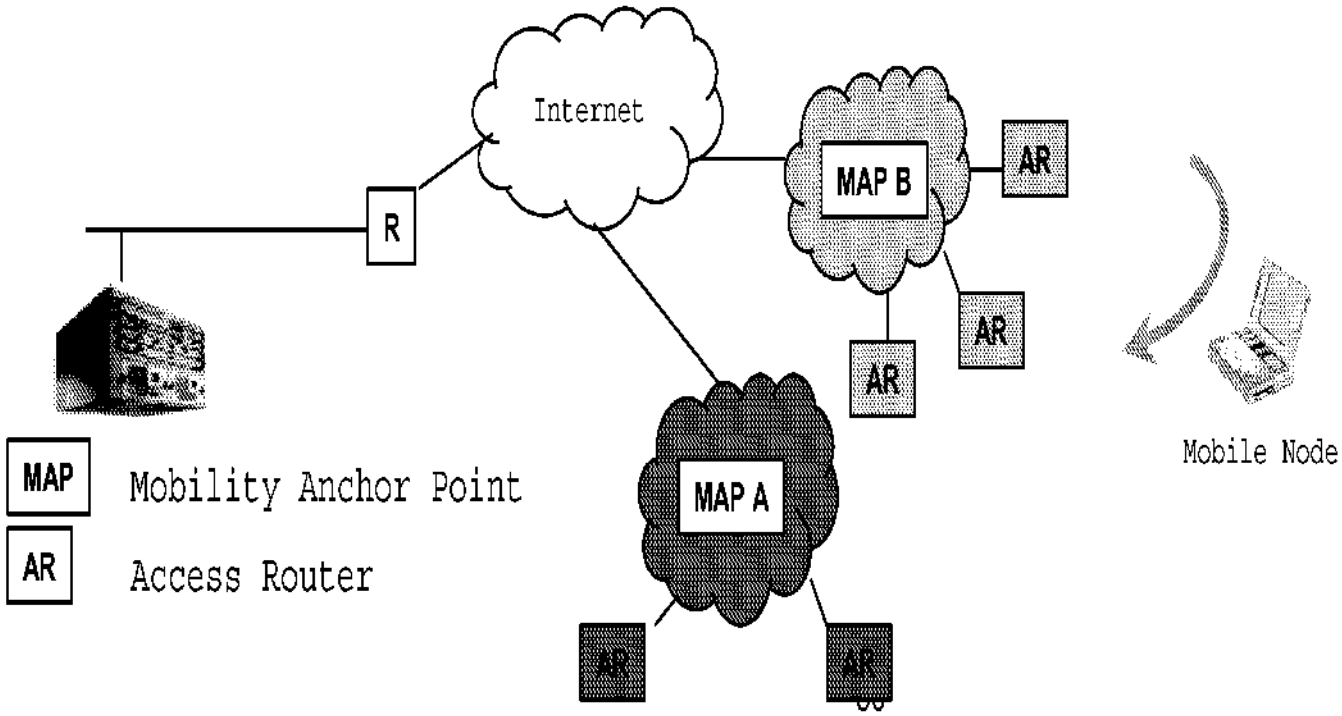
- Extension to Mobile IPv6
- Introduces hierarchical registration scheme

Scalability

- Not always registration to Home Agent necessary

Handoff

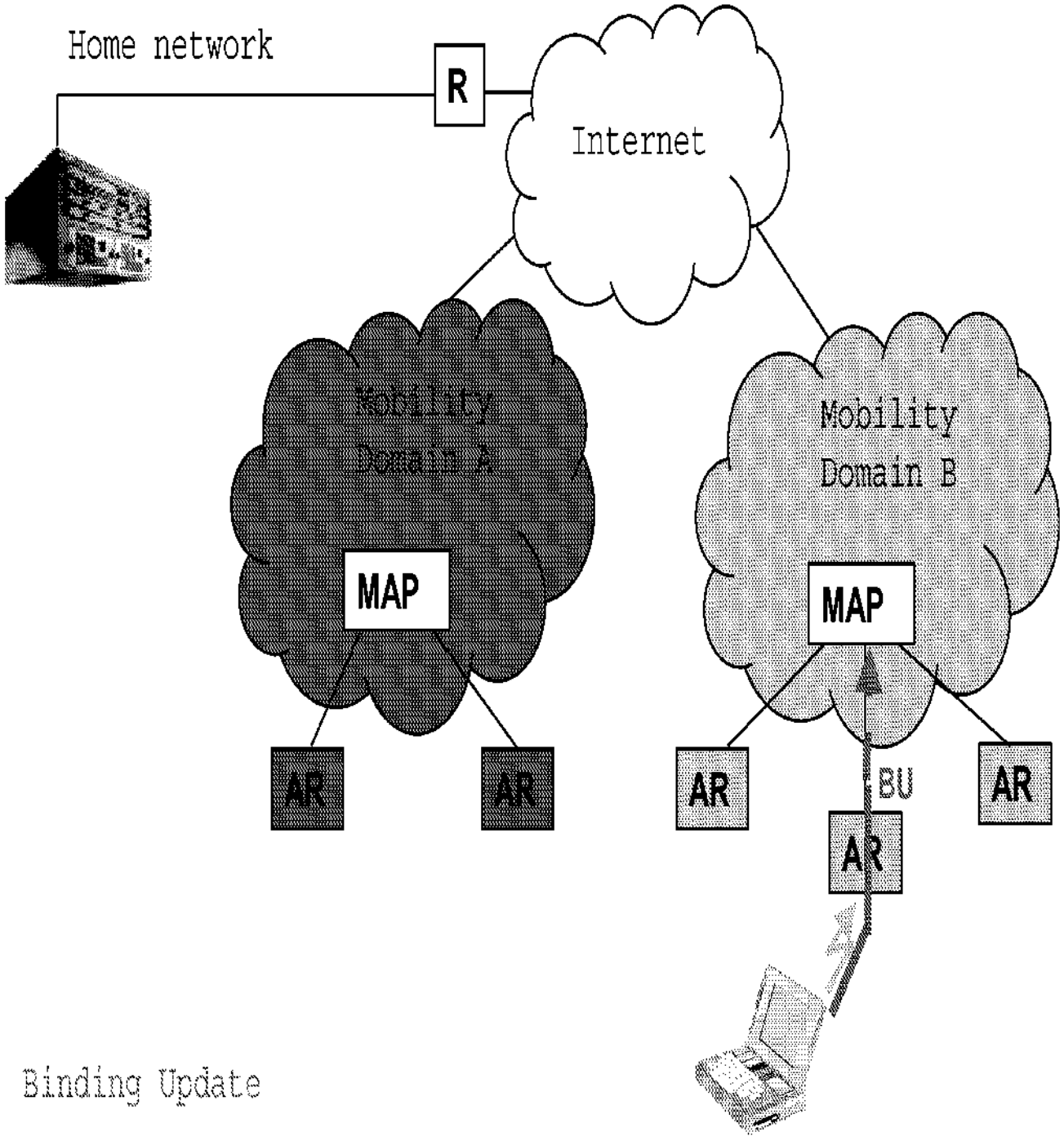
- Local registration decreases Handoff delay



WPI

RED HAT 1010
RED HAT V. CAS

Example 1: Mobility within Domain

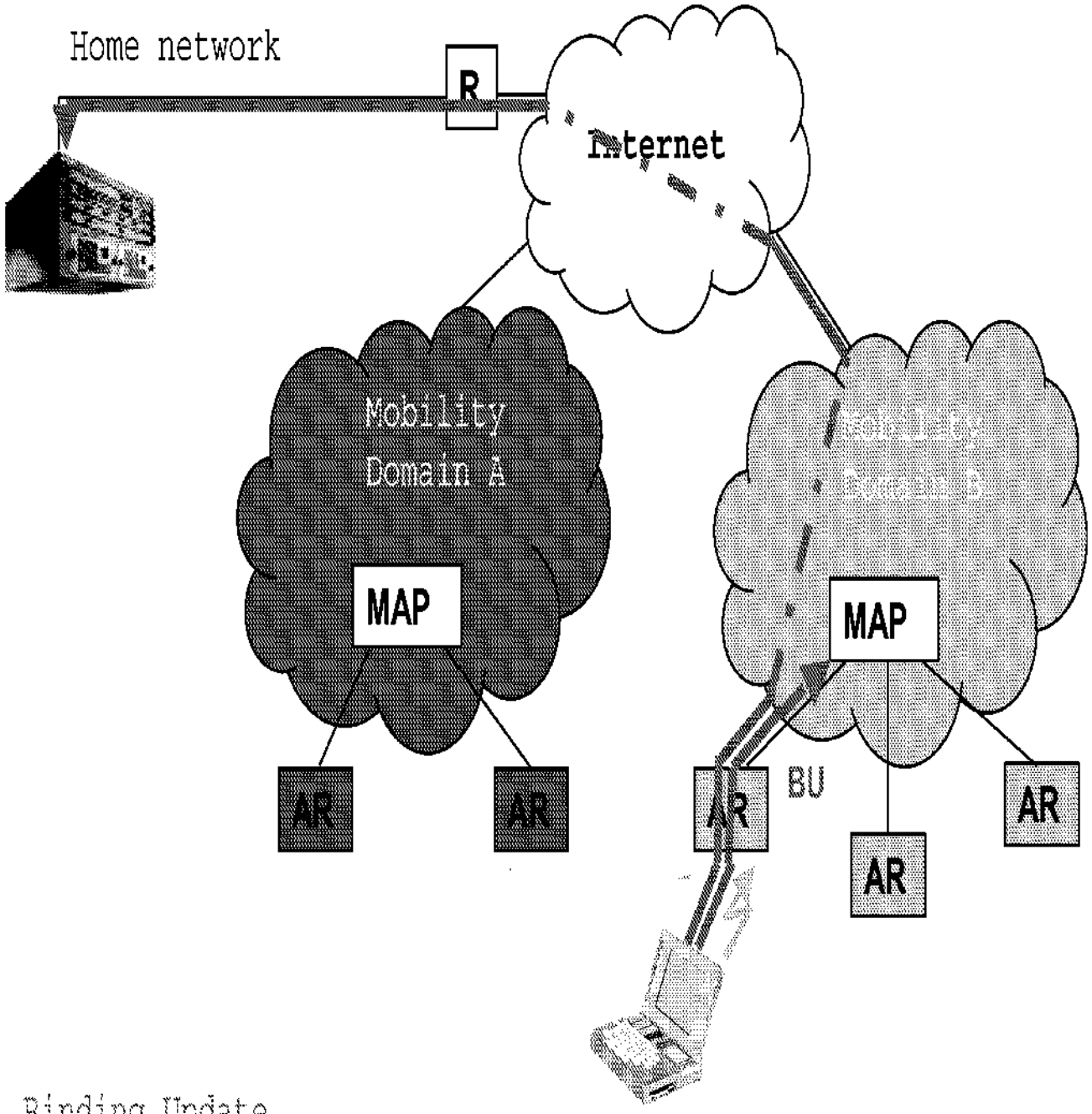


BU Binding Update

WPI

RED HAT 1010
RED HAT V. CAS

Example 2: Mobility between Domains



WPI

RED HAT 1010
RED HAT V. CAS

Summary

Both “sides”, Internet and Cellular Communication, have recognized the promising potential of the Mobile Internet market

Co-operation between organizations of the Internet and Cellular Communication side are established

IPv6 and Mobile IPv6 are seen as an efficient and scalable solution for the future Mobile Internet

Numerous research activities take place in the area of IPv6 for mobile users

From the technical side not all problems are solved now - but we are doing a good job here



Diversity of today's available mobile devices



WPI

Q & A

Thanks
for your attention!

WPI



Electronic Acknowledgement Receipt

EFS ID:	22813644
Application Number:	14512414
International Application Number:	
Confirmation Number:	9530
Title of Invention:	MULTIPATH COMMUNICATION DEVICES AND METHODS
First Named Inventor/Applicant Name:	Eric M. Delangis
Customer Number:	64828
Filer:	Harold C. McGurk
Filer Authorized By:	
Attorney Docket Number:	ED 2000
Receipt Date:	02-JUL-2015
Filing Date:	11-OCT-2014
Time Stamp:	13:27:20
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Transmittal Letter	14512414_IDS_Transmittal_Letter.pdf	72541 21b3e49a912e27 1b813f0f22f3327abd70bc5cb2	no	1

Warnings:	
Information:	

RED HAT 1010
RED HAT V. CAS

2	Non Patent Literature	cluster-optim.pdf	244920 776ee3336d847d252d15aff89b976c740966 0edfd	no	16
Warnings:					
Information:					
3	Non Patent Literature	draft-huitema-multi-homed-01.pdf	29967 574575ae83a9502ac0062791 76adc07c326c 45e6	no	18
Warnings:					
Information:					
4	Non Patent Literature	e2emobility.pdf	129661 4c2211f881980326a9b2496f9f07d875eff8e 56d3	no	12
Warnings:					
Information:					
5	Non Patent Literature	mysore97new.pdf	229989 542e7070c928df8cc78eae84ddba5cb33fc2 602a7	no	12
Warnings:					
Information:					
6	Non Patent Literature	rfc793.pdf	156548 25109bfbfcf3f7562123 5bb619299b99b87e bf2f	no	90
Warnings:					
Information:					
7	Non Patent Literature	rfc1323.pdf	96387 bba025c7f944b9526453270c12f696edf38 b1b4	no	37
Warnings:					
Information:					
8	Non Patent Literature	rfc2002.pdf	227685 06639aaa669a9f9d0f318b22b5af13ac08d4 305d	no	79
Warnings:					
Information:					
9	Non Patent Literature	rfc3012.pdf	59529 60ac1b9990d1bcf18e6dbf32e5f90fc46d79ae f8bd	no	17
Warnings:					
Information:					
10	Information Disclosure Statement (IDS) Form (SB08)	14512414_IDS_7-2-2015.pdf	612737 018b076bff1f28e9d3b46bf2accacf848a18 d62d	no	5
Warnings:					
Information:					

A U.S. Patent Number Citation or a U.S. Publication Number Citation is required in the Information Disclosure Statement (IDS) form for autoloading of data into USPTO systems. You may remove the form to add the required data in order to correct the Informational Message if you are citing U.S. References. If you chose not to include U.S. References, the image of the form will be processed and be made available within the Image File Wrapper (IFW) system. However, no data will be extracted from this form. Any additional data such as Foreign Patent Documents or Non Patent Literature will be manually reviewed and keyed into USPTO systems.

11	Non Patent Literature	Mobile_IPv6_fonts.pdf	6856520 41e580999c33742f2f9757c6a9f2d28f172 ff	no	56

Warnings:

Information:

Total Files Size (in bytes):	8716484

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

RED HAT 1010
RED HAT V. CAS

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION NO.:	**14/512,414**	EXAMINER:	
INVENTOR:	Eric M. Delangis	GROUP ART UNIT:	2652
FILED:	OCTOBER 11, 2014	CONFIRMATION NO.:	9530
TITLE:	Multipath Communication Devices and Methods		
ATTORNEY NO:	ED 2000		

Commissioner For Patents
PO Box 1450
Alexandria, VA 22313-1450

INFORMATION DISCLOSURE STATEMENT TRANSMITTAL LETTER

This submission does not represent that a search has been made or that no better art exists and does not constitute an admission that each or all of the listed documents are material or constitute prior art. If the Examiner applies any of the documents as prior art against any of the claims in the present application, all rights are reserved to take appropriate action to establish the patentability of the disclosed invention over the present application.

Best Regards,

Date: July 2, 2015

H C McGurk IV, Reg. No. 34,964
Attorney for Applicant

The Law Office of Clay McGurk
PO Box 1488
Orange, CA 92856
Phone: **949.395.3942**

CERTIFICATE OF ELECTRONIC (EFS-WEB) TRANSMISSION

I hereby certify that this correspondence is being transmitted via the U.S. Patent Office's electronic filing system in accordance with 37 C.F.R. § 1.8(a)(1) from the **Pacific Time Zone** of the United States on the local date shown below.

Dated: July 2, 2015

By: Clay McGurk

RED HAT 1010
RED HAT V. CAS

Cluster Computing 1 (1998) 161–176

Invited paper

Route Optimization for Mobile IP

Charles E. Perkins[a] and David B. Johnson[b]

[a] *Sun Microsystems, Menlo Park, CA 94025, USA*
[b] *Carnegie Mellon University, Pittsburgh, PA 15213-3891, USA*

Route Optimization has been designed within the IETF to ameliorate the problem of triangle routing, a routing artifact introduced by Mobile IP's requirement to route packets destined for a mobile node by way of its home network. In this article, we describe the current protocol specification for the Route Optimization protocol, concentrating on design decisions and justifications. Once the basic mechanisms are explained, we show how they are applied to enable foreign agents to offer smooth handoffs for mobile nodes, and describe the security operations that enable reliable operation of this handoff between foreign agents with which a mobile node has no pre-existing security relationship.

1. Introduction

Mobile IP has been developed in the Internet Engineering Task Force (IETF) as a solution to provide nearly transparent roaming for IP-addressable mobile nodes [15]. Physical effects such as variable bandwidth make it impossible for roaming to be completely transparent to applications, but such effects are forgivable since they are inevitable. Other problems such as routing anomalies and faulty congestion control are more difficult to understand, more difficult to diagnose, and less likely to be tolerated by typical users. Route Optimization is an attempt to solve the former problem, by reducing or eliminating the routing anomalies introduced by the base Mobile IP specification.

In this paper, we define the problem and give the details of one possible approach towards a solution. As with any routing problem, a robust solution needs to incorporate good security techniques, to avoid the possibility that a malicious network entity might introduce fraudulent routing information and thus disrupt communications. Not solving the problem is better than solving the problem in a way that offers opportunities for corrupting the integrity of the routing tables of computers which need to communicate with mobile nodes.

Section 2 of this paper gives a brief overview of Mobile IP and the *Route Optimization* extensions to it, describing their component parts and design. Section 3 describes the process of *binding cache* maintenance and details each of the messages used by Route Optimization for this purpose. The application of this cache maintenance for providing *smooth handoffs* as a mobile node moves from one foreign agent to the next is presented in section 4, and section 5 then presents the problem of the managing mobility security associations needed by Route Optimization. Accomplishing smooth handoffs requires the use of *registration keys*, and the proposed methods for establishing those keys are detailed in section 6, with the message formats and processing steps detailed in sections 7–9. To further illustrate the power of Route Optimization techniques, in section 10 we show how binding cache maintenance has been applied in the protocol design for mobility support in IPv6, the new version of IP now being designed in the IETF. Lastly, in section 11, we summarize and present conclusions.

2. Overview

The development of powerful laptop computers with medium-speed wireless communications adapters has been the driving motivation for the creation of new protocols. Mobile IP [15] enables a mobile node to move from place to place in the Internet, maintaining active connections and presenting to typical Internet nodes the illusion that it remains present on its *home network*. Communications with the mobile node proceed by use of the mobile node's *home address*, which does not depend on the node's current point of attachment to the Internet. At each point of attachment, the mobile node acquires a *care-of address*, which it must report to its *home agent* on its home network by a process called *registration*.

In all situations considered in the rest of this paper, the mobile node obtains the care-of address by interaction with a *foreign agent* on the *foreign network*. The foreign agent offers the care-of address using ICMP [17] by including it as part of a specially modified *Router Advertisement* [3], which is then known as an *Agent Advertisement*. The Agent Advertisement also includes a maximum time duration, or *lifetime*, for which the mobile node may consider the care-of address valid, and which bounds the lifetime permissible within the mobile node's Registration Request that is subsequently transmitted to the home agent. When a mobile node which is using a care-of address detects that it is no longer receiving the Agent Advertisements from its current foreign agent, or in some other way detects that it no longer is in contact with that foreign agent, it assumes that the care-of address is no longer valid. The mobile node then begins to search for a new care-of address, presumably from another foreign agent.

© Baltzer Science Publishers BV

RED HAT 1010
RED HAT V. CAS

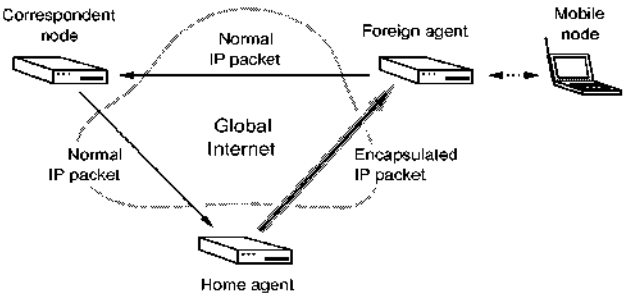


Figure 1. Overview of the base Mobile IP protocol.

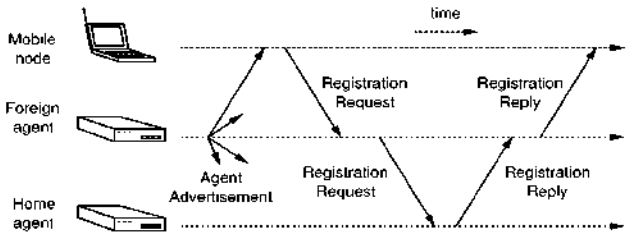


Figure 2. Overview of the Mobile IP registration process.

Figure 1 shows an overview of the base Mobile IP protocol, including the relative placement of Mobile IP agents and networks. A *correspondent node* is any IP host or router that communicates with a mobile node. A correspondent node, itself, may be either mobile or stationary. To send a packet to a mobile node, a correspondent node transmits the packet to the mobile node's home address, which causes the packet to be routed toward the mobile node's home network. There, the packet is intercepted by the mobile node's home agent. The home agent then tunnels the packet to the mobile node's current foreign agent, using the care-of address as the tunnel destination. The foreign agent decapsulates the packet and delivers it locally to the mobile node. If a mobile node sends a packet to a correspondent node, it simply sends it in the same way as if it were at home, but uses its foreign agent as the default router for delivering the packet. The foreign agent, simply acting as a router, then forwards the packet directly to the correspondent node.

The processes of agent advertisement and registration of a care-of address are illustrated in figure 2. Whenever a mobile node registers, its home agent (if it approves the request) associates the mobile node's home address with the care-of address and lifetime together in a routing record known as a *binding*. The home agent thus maintains a *binding cache* containing all the bindings for all those mobile nodes that are using its services while they are away from the home network. The home agent also performs whatever functions are necessary [15] to manage the interception of all packets destined for its registered mobile nodes. Once the home agent has intercepted a packet for a mobile node, it consults its binding cache to *tunnel* (encapsulate) the packet, setting the tunnel destination to be the mobile node's care-of address [13,14]. This tunneling process appears as additional routing overhead on all packets addressed to the mobile node.

This description of packet delivery to and from a mobile node illustrates a routing anomaly caused by the operation of Mobile IP. Packets sent *to* a mobile node are routed indirectly through the mobile node's home agent and are then tunneled to the mobile node, whereas packets sent *from* a mobile node are routed directly to the correspondent node. Round trip communications thus travel along three distinct routing paths, and thus this routing anomaly is known as *triangle routing*. This use of triangle routing can be seen in figure 1. All of these operations are specified by the base Mobile IP standard [15].

For the correspondent node to eliminate triangle routing, it must have some information regarding the current location of the mobile node – in other words, the correspondent node will have to discover the mobile node's care-of address, and to maintain a binding cache giving this care-of address for use in tunneling its own packets to the mobile node's care-of address.

Whenever the care-of address later changes, the correspondent node must update its binding cache. Route Optimization accomplishes this by sending a *Binding Update* to the correspondent node. The production and consumption of these Binding Updates form the heart of the operation of the Route Optimization protocol. And, just as the home agent cannot accept an unauthenticated Registration Request from a mobile node, neither can a correspondent node accept an unauthenticated Binding Update. One of the guiding principles for deciding how to send Binding Updates to correspondent nodes is that the correspondent node must be able to trust the sender. In the current Internet, deployment of authentication technology and key distribution is typically handled interactively, and thus not so amenable to the kind of automatic operation required for Route Optimization. Thus, our proposal has been designed to reduce the number of mobility agents which have to maintain trust with correspondent nodes.

When a mobile node moves to a new care-of address, any existing binding cache entries for the mobile node in different correspondent nodes' binding caches become out-of-date. An out-of-date binding causes a correspondent node to tunnel packets to an old care-of address. Unfortunately, in base Mobile IP, such packets are likely to get dropped because the mobile node's previous foreign agent is not notified of the mobile node's movement to its new care-of address. On the other hand, if the previous foreign agent can maintain a binding cache entry for a mobile node which had previously been visiting, that foreign agent can deliver such misdirected packets to the mobile node's current care-of address. The Binding Update messages used to inform correspondent nodes about a mobile node's care-of address are also used to inform the mobile node's previous foreign agent when the mobile node moves and acquires a new care-of address. Then, the previous foreign agent can act as a temporary forwarder for traffic destined to the mo-

RED HAT 1010
RED HAT V. CAS

bile node, until all of the relevant correspondent nodes have updated their binding cache entries for the mobile node. This process is called *smooth handoff*, and is described in detail in section 4.

While using the same messages, smooth handoff introduces a new requirement for establishing trust between a mobile node and each of its foreign agents. In base Mobile IP, and for the purposes of reliably delivering Binding Updates to correspondent nodes, the foreign agents are largely passive agents. No particular trust model is assumed to be in place between the mobile node and the foreign agent. For smooth handoff to work correctly, the previous foreign agents have to be able to authenticate Binding Updates, which means they must share a secret with the node transmitting the binding update. Moreover, to correctly handle packets in flight, the previous foreign agent must get the binding updates as soon as possible. Using the model of a mobile node moving from one wireless access point to the next nearby one, with each access point serving as or for a foreign agent, we choose to enable the new foreign agent to deliver the necessary Binding Update to the previous foreign agent as soon as possible. Establishing a way for the foreign agent to trust the mobile node, while difficult, is possible using any one of various techniques detailed later in the paper. All that the previous foreign agent has to know is that the Binding Update comes from the same mobile node as had been registered with it; this is accomplished by exchanging key information as part of the previous Registration Request and Registration Reply.

As a matter of terminology, the *mobility security associations* between nodes in this paper are usually indicated by a *Security Parameters Index*, or SPI. The SPI indicates which of possibly many security associations between two nodes is to be used when performing the necessary security operations. For instance, an SPI may indicate that authentication is to be performed by comparing the results of an MD5 computation operating on a stream of data, with some secret information included both before and after the data to be authenticated. SPIs are an important field in the registration key messages and other messages relating to a binding with the mobile node's new care-of address. SPIs are selected arbitrarily from the available 32-bit unsigned numbers, except that SPI numbers 0 through 255 are reserved and not allowed to be used in any mobility security association.

3. Binding cache maintenance messages

A correspondent node may create or update a binding cache entry for a mobile node only when it has received and authenticated the mobile node's mobility binding. In addition, a node may use any reasonable strategy for managing the space within its binding cache. When a new entry needs to be added to the binding cache, the node may choose to drop any entry already in the cache, if needed, to make space for the new entry. For example, a *least-recently used* (LRU) strategy for cache entry replacement is likely to work well. This is in contrast to the way that a home agent should manage the registrations for mobile nodes registered with it. A home agent should not drop a registration until the expiration of the lifetime of the binding established during the registration process.

When sending an IP packet, if the sending node has a binding cache entry for the destination node, it should tunnel the packet to the mobile node's care-of address using the encapsulation techniques used by home agents [7,8,13,14].

When a mobile node's home agent intercepts a packet from the home network and tunnels it to the mobile node, as shown in figure 1, the home agent may deduce that the original source of the packet has no binding cache entry for the destination mobile node. The home agent should then send a Binding Update message to the original source node, informing it of the mobile node's current mobility binding. No acknowledgement for such a Binding Update message is needed, since additional future packets from this source node intercepted by the home agent for the mobile node would cause transmission of another Binding Update. For a Binding Update to be authenticated by the original source node, the source node and the home agent must have established a mobility security association.

Similarly, when a foreign agent receives a tunneled packet, if it has a binding cache entry for the destination node, indicating that the mobile node has established a binding elsewhere (not with this foreign agent), the foreign agent may deduce that the tunneling node has an out-of-date binding cache entry for the mobile node. In this case, the foreign agent should send a *Binding Warning* message to the mobile node's home agent, advising it to send a Binding Update message to the node that tunneled this packet. The mobile node's home agent can be determined from the binding cache entry, because the home agent's address is learned from the Binding Update that established the cache entry. The address of the node that tunneled the misdirected packet can be determined from the packet's header, since the address of the node tunneling this packet is the *outer* source address of the encapsulated packet. As in the case of a Binding Update sent by the mobile node's home agent, no acknowledgement of this Binding Warning is needed, since additional future packets for the mobile node tunneled by the same node will cause the transmission of another Binding Warning.

In addition to the Binding Update and Binding Warning messages, Route Optimization makes use of two additional types of messages for binding cache maintenance: Binding Request and Binding Acknowledgement. These four message types are distinguished by a one-octet *type* field, with the message type values chosen from the numbering space defined in the base Mobile IP specification for messages sent to UDP port 434.

The following sections describe in detail each of the Route Optimization messages used for binding cache maintenance, as well as a message extension used by them for authentication. In addition, Route Optimization requires

RED HAT 1010
RED HAT V. CAS

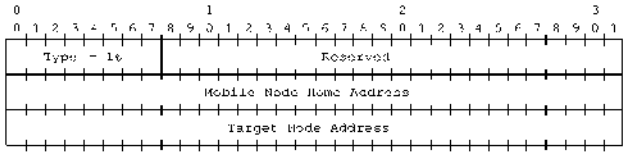


Figure 3. Binding Warning message format.

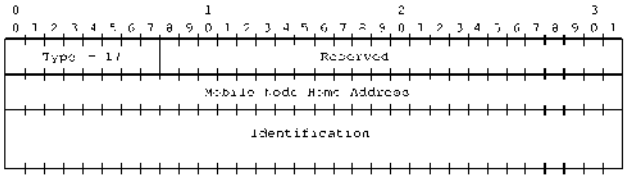


Figure 4. Binding Request message format.

one minor change to the existing Mobile IP Registration Request message: a new flag bit must be added, replacing a previously unused, reserved bit in the message. This modified Registration Request message format is also described below. In the description of each message format, any field marked as *reserved* must be set to all zeroes when transmitted and must be ignored upon reception.

3.1. Binding Warning message

A *Binding Warning* message is used to advise a mobile node's home agent that another node appears to have either no binding cache entry or an out-of-date binding cache entry for some mobile node. When any node receives and decapsulates a tunneled packet for which it is not the current foreign agent for the destination mobile node, if it forwards the packet to a new care-of address based on an entry in its own binding cache, it should send a Binding Warning message to the mobile node's home agent indicated in that binding cache entry. No authentication of the Binding Warning message is necessary, since it does not directly affect the routing of IP packets to the mobile node.

The format of the Binding Warning message is illustrated in figure 3. The mobile node's home IP address is followed by the *target node address*, which is the IP address of the node that tunneled the packet causing the Binding Warning message. A mobile node's home agent will thus receive a Binding Warning message if a node maintaining a binding cache entry for one of the home agent's mobile nodes uses an out-of-date entry. When a home agent receives a Binding Warning message, it should send a Binding Update message including the mobile node's current binding to the target node.

A foreign agent must provide some mechanism to limit the rate at which it sends Binding Warning messages to the same node about any given mobility binding. This rate limiting is especially important because it is expected that, within the short term, many Internet nodes will not support maintenance of a binding cache. In this case, continual transmissions of Binding Warning messages will only waste processing resources at the foreign agent, home agent, and correspondent node, and along the Internet path between these nodes.

3.2. Binding Request message

A *Binding Request* message is used by a node to request a mobile node's current mobility binding from the mobile node's home agent. A node wanting to provide continued service with a particular binding cache entry may attempt to reconfirm that mobility binding before the expiration of the registration lifetime. Such reconfirmation of a binding cache entry may be appropriate when the node has indications (such as an open transport-level connection to the mobile node) that the binding cache entry is still needed. This reconfirmation is performed when the node sends a *Binding Request* message to the mobile node's home agent, requesting a new Binding Update message with the mobile node's current mobility binding. The node maintaining the binding cache entry should also keep track of the home agent's address, to be able to fill in the destination IP address of future Binding Requests.

The format of the Binding Request message is illustrated in figure 4, and contains the home address of the mobile node to which the Binding Request refers. Following that is a 64-bit sequence number, assigned by the node sending the Binding Request message, which is used to assist in matching requests with replies and in protecting against replay attacks.

When the home agent receives a Binding Request message, it consults its list of registered mobile nodes and determines the correct binding information to be sent to the requesting node. Before satisfying the request, the home agent must check whether or not the mobile node has allowed the information to be disseminated. If the mobile node specified the *private* (P) bit in its Registration Request message (as shown in section 3.6), then the home agent returns a Binding Update in which both the care-of address is set equal to the mobile node's home address and the lifetime is set to zero. Such a Binding Update message indicates that the binding cache entry for the specified mobile node should be deleted.

3.3. Binding Update message

The *Binding Update* message is used for notification of a mobile node's current mobility binding. It should be sent by the mobile node's home agent in response to a Binding Request message or a Binding Warning message. It should also be sent by a mobile node, or by the foreign agent with which the mobile node is registering, when notifying the mobile node's previous foreign agent that the mobile node has moved.

The format of the Binding Update message is illustrated in figure 5, and contains the mobile node's home address and care-of address. Setting the care-of address in the Binding Update equal to the home address of the mobile node

RED HAT 1010
RED HAT V. CAS

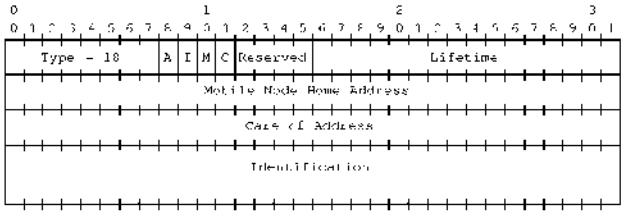


Figure 5. Binding Update message format.

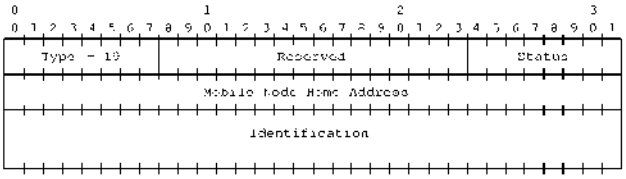


Figure 6. Binding Acknowledgement message format.

means that any existing binding cache entry (and visitor list entry, in the case of a mobile node's previous foreign agent) for the mobile node should be deleted. The Binding Update also contains a *lifetime* for this binding, and any binding cache entry created or modified as a result of receipt of this Binding Update must be deleted once this lifetime expires. A value of all ones for the lifetime indicates infinity; a value of zero means the same thing as setting the care-of address equal to the home address. When sending the Binding Update message, the home agent should set this lifetime to the remaining registration lifetime.

The Binding Update message also contains four defined flag bits. The *acknowledge* (A) bit is set by the node sending the Binding Update message to request a Binding Acknowledgement message be returned. The *identification present* (I) bit is set to indicate that the *identification* field is present in the message. The *minimal encapsulation* (M) and *Generic Routing Encapsulation* (GRE) (G) bits indicate support for receiving packets encapsulated with either of these two optional encapsulation protocols [7,14].

When a node receives a Binding Update message, it must verify the authentication in the message, using the mobility security association it shares with the mobile node's home agent. The authentication data is found in the *Route Optimization Authentication extension* (section 3.5), which must be present in the message. If the authentication succeeds, then a binding cache entry should be updated for use in sending future packets to the mobile node. Otherwise, an authentication exception should be raised.

As with the sending of Binding Warning messages, a home agent must provide some mechanism to limit the rate at which it sends Binding Update messages to to the same node about any given mobility binding. Since in the short term, some nodes may not support maintenance of a binding cache, continual transmissions of Binding Update messages will only waste processing resources at the home agent and correspondent node, and along the Internet path between these nodes.

The 64-bit *identification* field is present unless the Binding Update is sent as part of a *smooth handoff*. When using nonces for replay protection [15], the identification field in the Binding Update message is used slightly differently than when timestamps are used [15], to still allow replay protection even though the Binding Update is not being sent in reply to a request directly from the target node. In this case, the home agent is required to set the high-order 32 bits of the identification field to the value of the nonce that will be used by the home agent in the next Binding Update message sent to this node. The low-order 32 bits of the identification field must be set to the value of the nonce being used for this message.

Thus, on each Binding Update message, the home agent communicates to the target node the value of the nonce that will be used next time, and if no Binding Updates are lost in the network, the home agent and the target node can remain synchronized with respect to the nonces being used. If, however, the target node receives a Binding Update with what it believes to be an incorrect nonce, it may resynchronize with the home agent by using a Binding Request message.

3.4. Binding Acknowledgement message

A *Binding Acknowledgement* message is used to acknowledge receipt of a Binding Update message. It should be sent by a node receiving a Binding Update message if the *acknowledge* (A) bit is set in the Binding Update.

The format of the Binding Acknowledgement message is illustrated in figure 6, and contains a *status* field in addition to the mobile node's home address and the usual identification field for replay protection, copied from the Binding Update message.

If the status is nonzero, the Acknowledgement is negative. For instance, if the Binding Update was not accepted, but the incoming packet has the *acknowledge* flag bit set, then the status code should be set appropriately in the Binding Acknowledgement message. Status values are defined to indicate different conditions for which the Binding Update may have been rejected, including administrative prohibition, lack of sufficient resources, authentication failure, or a mismatch in the expected identification value.

3.5. Route Optimization Authentication extension

The *Route Optimization Authentication* extension is used to authenticate Binding Update and Binding Acknowledgement messages. It has the same format and default algorithm support requirements as the three authentication extensions defined for base Mobile IP [15], but is distinguished by its message type value. The authenticator value is computed from the stream of bytes including the shared secret, the UDP payload, all prior extensions in their entirety (that is, the Route Optimization management message), and the type and length of this extension, but not including the authenticator field itself nor the UDP header.

RED HAT 1010
RED HAT V. CAS

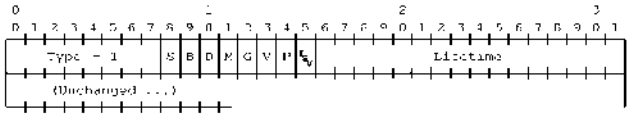


Figure 7. Modified Registration Request message format.

For implementations that can support more than the mandatory base authentication algorithm, other optional authentication algorithms such as the more secure HMAC authenticator [11] could also be used if specified in the mobility security association.

3.6. Modified Registration Request message

The format of the modified *Registration Request* message is illustrated in figure 7. It contains one new bit, the *private* (P) bit, enabling the mobile node to indicate that it would like its home agent to keep its mobility binding *private*. Normally, the home agent sends Binding Update messages to correspondent nodes as needed to allow them to cache the mobile node's binding. If the mobile node sets the *private* bit in the Registration Request message, the home agent is not allowed to send the mobile node's binding in any Binding Update message. Instead, each Binding Update message should give the mobile node's care-of address equal to its home address, and should give a lifetime value of 0.

4. Smooth handoffs

This section provides a description of the proposed operation of smooth handoff from a mobile node's *previous* foreign agent to its *new* foreign agent when the mobile node initiates a new registration.

4.1. Smooth handoff overview

When a mobile node moves and registers with a new foreign agent, the base Mobile IP protocol does not notify the mobile node's previous foreign agent. IP packets intercepted by the home agent after the new registration are tunneled to the mobile node's new care-of address, but packets in flight that had already been intercepted by the home agent and tunneled to the old care-of address when the mobile node moved are likely to be lost and are assumed to be retransmitted by higher-level protocols if needed. The old foreign agent eventually deletes its visitor list entry for the mobile node after the expiration of the registration lifetime.

Route Optimization provides a means for the mobile node's previous foreign agent to be reliably notified of the mobile node's new mobility binding, allowing packets in flight to the mobile node's *previous* care-of address to be forwarded to its *new* care-of address. Any packets tunneled to the mobile node's previous foreign agent, from correspondent nodes with out-of-date binding cache

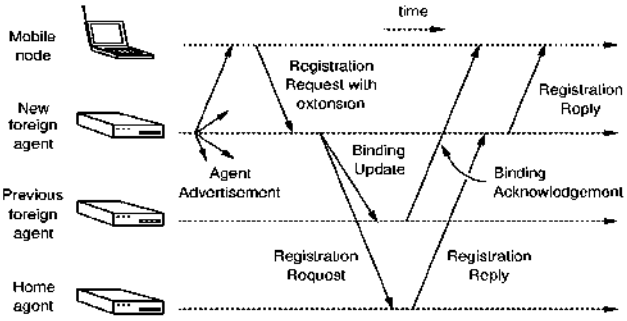


Figure 8. Smooth handoff.

entries for the mobile node, can also be recovered in this way. Finally, this notification allows any resources consumed by the mobile node at the previous foreign agent (such as an allocated radio channel) to be released immediately, rather than waiting for its registration lifetime to expire.

The mobile node requests that its new foreign agent attempt to notify its previous foreign agent on its behalf by including a *Previous Foreign Agent Notification* extension in its Registration Request message sent to the new foreign agent. The new foreign agent then builds a Binding Update message and transmits it to the mobile node's previous foreign agent during registration, requesting an acknowledgement from the previous foreign agent. The notification will typically include the mobile node's new care-of address, allowing the previous foreign agent to create a binding cache entry for the mobile node to serve as a *forwarding pointer* [9] to its new location. This process is illustrated in figure 8. The figure also shows how the binding cache maintenance messages are sent during the Mobile IP registration process. Any tunneled packets for the mobile node that arrive at its previous foreign agent after the forwarding pointer has been created can then be re-tunneled to the mobile node's new care-of address.

For this smooth handoff to be secure, during registration with a new foreign agent, the mobile node and the previous foreign agent must have a security association. The security association is used to authenticate the notification sent to the previous foreign agent.

The mobile node is responsible for occasionally retransmitting a Binding Update message to its previous foreign agent until the matching Binding Acknowledgement message is received, or until the mobile node can be sure that the foreign agent has expired its binding. The mobile node is likely to select a small timeout value for the lifetime available to such bindings sent to previous foreign agents.

4.2. Previous Foreign Agent Notification extension

The format of the Previous Foreign Agent Notification extension is illustrated in figure 9. The Previous Foreign Agent Notification extension contains only those values

RED HAT 1010
RED HAT V. CAS

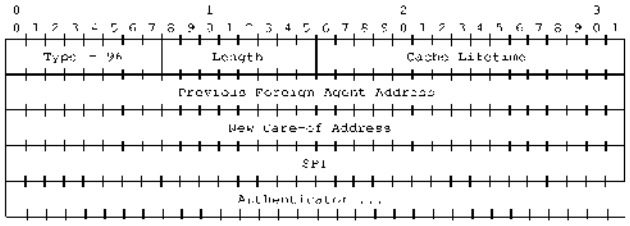


Figure 9. Previous Foreign Agent extension format.

not otherwise already contained in the Registration Request message that are needed for the new foreign agent to construct the Binding Update message. It contains the mobile node's new care-of address (or, sometimes, the mobile node's home address), the previous foreign agent's address, and sufficient information for the new foreign agent to build the expected Binding Update for the previous foreign agent. If the mobile node just wants the previous foreign agent to forget entirely about the mobile node, it uses its home address in the notification extension.

The cache lifetime is copied into the lifetime field of the Binding Update message, and is the number of seconds remaining before the binding cache entry created by the previous foreign agent must be considered expired. A value of zero indicates that the previous foreign agent should not create a binding cache entry for the mobile node once it has deleted the mobile node's registration in its visitor list.

When building the Binding Update, the new foreign agent must create a *Route Optimization Authentication* extension. The *SPI* and authenticator value are copied into the authentication extension by the new foreign agent. This authenticator is calculated by the mobile node only over the predicted Binding Update message body, using a security association shared with its previous foreign agent. The SPI used for the authentication calculation can be one created as part of the establishment of the registration key by one of the methods in section 7.

4.3. Smooth handoffs and Binding Acknowledgements

When a foreign agent receives a *Previous Foreign Agent Notification* message, it creates a *Binding Update* for the previous foreign agent, using the specified SPI and precomputed authenticator sent to it by the mobile node. The Binding Update message is also required to set the *acknowledge* bit, so that the previous foreign agent will know to send a *Binding Acknowledgement* message back to the mobile node.

When the previous foreign agent receives the Binding Update message, it will authenticate the message using the mobility security association indicated by the SPI. If the message authentication is correct, the visitor list entry for this mobile node at the previous foreign agent will be deleted and a Binding Acknowledgement message returned to the sender. In addition, if a new care-of address was included in the Binding Update message, the previous foreign agent will create a binding cache entry for the mobile node. The previous foreign agent can then tunnel packets to the mobile node's new care-of address using that binding cache entry, just as any node maintaining a binding cache.

This Binding Acknowledgement returned by the previous foreign agent is addressed to the mobile node, and thus must be tunneled using the new binding cache entry. The tunneled acknowledgement then should be delivered directly to the new foreign agent, without having to go to the home network. This creates an interesting problem for the new foreign agent when it receives the acknowledgement before the Registration Reply from the home agent. It is suggested that the new foreign agent deliver the acknowledgement to the mobile node anyway, even though the mobile node is technically unregistered. If there is concern that this provides a loophole for unauthorized traffic to the mobile node, the new foreign agent could limit the number of packets delivered to the unregistered mobile node to this single instance. Alternatively, a new extension to the Registration Reply message can be defined to carry along the acknowledgement from the previous foreign agent. This latter approach would have the benefit that fewer packets would be transmitted over bandwidth-constrained wireless media during registration.

When the Binding Acknowledgement message from the previous foreign agent is received by the new foreign agent, it decapsulates it and sends it to the mobile node. In this way, the mobile node can discover that its previous foreign agent has received the Binding Update message. The mobile node must be certain that its previous foreign agent has been notified about its new care-of address, because otherwise the previous foreign agent could become a "black hole" for packets destined for the mobile node based on out-of-date binding cache entries at other nodes. The new foreign agent has no further responsibility for helping to update the binding cache at the previous foreign agent, and does not retransmit the message even if no acknowledgement is received.

If the acknowledgement has not been received after sufficient time, the mobile node is responsible for retransmitting another Binding Update message to its previous foreign agent. Although the previous foreign agent may have already received and processed the Binding Update message (the Binding Acknowledgement message may have been lost in transit to the new foreign agent), the mobile node should continue to retransmit its Binding Update message until the previous foreign agent responds with a Binding Acknowledgement.

It is possible that the binding cache entry created by the previous foreign agent from the information in the Binding Update from the new foreign agent will be deleted from its cache at any time. In this case, the previous foreign agent will be unable to re-tunnel subsequently arriving tunneled packets for the mobile node, and would resort to using a *special tunnel* [16]. Mobile nodes are expected to assign small lifetimes to such bindings so that they will not take

RED HAT 1010
RED HAT V. CAS

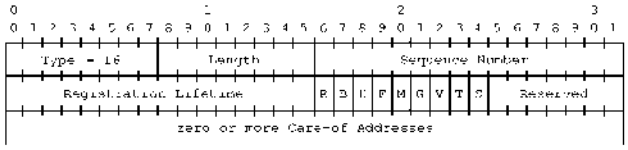


Figure 10. Modified Agent Advertisement message format.

up space in the foreign agent's binding cache for longer than necessary.

4.4. Modified Mobility Agent Advertisement extension

Performing smooth handoffs requires one minor change to the existing Mobile IP *Mobility Agent Advertisement* extension [15]. A new flag bit, the *smooth handoff* (S) bit, replaces a previously unused reserved bit in the extension, to indicate that the foreign agent supports smooth handoffs. By default, every foreign agent that supports smooth handoffs is expected to support at least the establishment of a registration key (see section 6) by using a *Diffie–Hellman* key exchange.

5. Mobility security association management

One of the most difficult aspects of Route Optimization for Mobile IP in the Internet today is that of providing authentication for all messages that affect the routing of packets to a mobile node. In the base Mobile IP protocol, only the home agent is aware of the mobile node's mobility binding and only the home agent tunnels packets to the mobile node. Thus, all routing of packets to the mobile node while away from its home network is controlled by the home agent. Authentication is currently achieved based on a manually established mobility security association between the home agent and the mobile node. Since the home agent and the mobile node are both owned by the same organization (both are assigned IP addresses within the same IP subnet), this manual configuration is manageable, and (for example) can be performed while the mobile node is at home.

However, with Route Optimization, authentication is more difficult to manage, since a Binding Update may in general need to be sent to almost any node in the Internet. Since no authentication or key distribution protocol is generally available in the Internet today, the Route Optimization procedures defined here may make use of the same type of manual key distribution as defined in the base Mobile IP protocol. For use with Route Optimization, a mobility security association held by a correspondent node or a foreign agent must include the same parameters (for instance, the style of replay protection) as required by base Mobile IP [15].

For a correspondent node to be able to create a binding cache entry for a mobile node, the correspondent node and the mobile node's home agent must have established a mobility security association. This mobility security association, though, could conceivably be used in creating and updating binding cache entries at this correspondent node for all mobile nodes served by this home agent. Doing so places the correspondent node in a fairly natural relationship with respect to the mobile nodes served by this home agent. For example, the mobile nodes may represent different people affiliated with the same organization owning the home agent, with which the user of the correspondent node often collaborates. The effort of establishing such a mobility security association with the relevant home agent may be more easily justified (as described below) than the effort of doing so with each individual mobile node. It is similarly possible for a home agent to have a manually established mobility security association with the foreign agents often used by its mobile nodes, or for a particular mobile node to have a manually established mobility security association with the foreign agents serving the foreign networks that it often visits.

In general, if the movement and communication patterns of a mobile node or the group of mobile nodes served by the same home agent are sufficient to justify establishing a mobility security association with the mobile node's home agent, users or network administrators are likely to do so. Rather than storing each mobility security association that it has established with many different correspondent nodes and foreign agents, a home agent may manage its mobility security associations so that each of them can be generated from a single *master key*. With the master key, the home agent could build a key for any given other node, e.g., by computing the node-specific key as

MD5(node-address || master-key || node-address)

where node-address is the IP address of the particular node for which the home agent is building a key, and master-key is the single master key held by the home agent for all mobility security associations it has established with correspondent nodes. The node-specific key is built by computing an MD5 hash over a string consisting of the master key with the node-address concatenated as a prefix and as a suffix.

Using this scheme, when establishing each mobility security association, the network administrator managing the home agent computes the node-specific key and communicates this key to the network administrator of the other node through some secure channel, perhaps by telephone. The mobility security association is configured at this other node in the same way as any mobility security association. At the home agent, though, no record need be kept that this key has been given out. The home agent need only be configured to know that this scheme is in use for all of its mobility security associations, or perhaps only for a specific set of its mobile nodes. When the home agent needs a mobility security association as part of Route Optimization, it builds the node-specific key based on the master key and the IP address of the other node with which it is attempting to authenticate.

RED HAT 1010
RED HAT V. CAS

6. Registration key establishment messages

When a mobile node registers with a foreign agent, it typically does not yet share a security association with the foreign agent. However, in order for the foreign agent to process future Binding Updates that it may receive, it must have such a security association. As described earlier, smooth handoffs rely on the Previous Foreign Agent Notification extension (section 4.2), which precedes the transmission of a Binding Update to the previous foreign agent created by the mobile node after it moves. Binding Updates provide the essential mechanism for accomplishing smooth handoffs between a previous foreign agent and a new foreign agent.

Foreign agents are expected to be cheap and widely available as Mobile IP becomes fully deployed. Mobile nodes will likely find it difficult to manage long-term security relationships with so many foreign agents. To securely perform the operations needed for smooth handoffs from one foreign agent to the next, however, any careful foreign agent should require assurance that it is getting authentic handoff information and is not arranging to forward in-flight packets to a forged destination. The registration key establishment messages are used with the Mobile IP Registration Request and Registration Reply messages to establish some trustworthy secret (and SPI) between a mobile node and its foreign agent when none exists beforehand, while allowing the use of fully trustworthy security associations between foreign agents and mobile nodes whenever they do exist.

In Mobile IP, the mobile node often cannot verify the identity of the foreign agent in any absolute terms. It can only act on the presumption that the foreign agent is performing its duties by correctly implementing the protocol. The exact identity of the foreign agent is not crucial to the process of establishing a registration key. Only an agreement to follow the protocol can be expected or enforced. If the mobile node has a way to obtain a certified public key for the foreign agent, then the identity may be established in a firmer fashion, but the needed public key infrastructure is not yet available in the Internet today. Therefore, the methods described here enable a mobile node to create a registration key with an *anonymous* foreign agent (i.e., one whose identity we may be unable to verify) during the registration process. Several methods for establishing a registration key have been proposed, and other methods of establishing keys may become available in the future, using an Internet public key infrastructure or Kerberos. Currently, the following methods have been proposed:

1. If the foreign agent and mobile node share a security association, it can be used to secure the Previous Foreign Agent Notification without need to establish a registration key.

2. If the home agent and foreign agent share a security association, the home agent can choose the new registration key.

3. If the foreign agent has a public key, it can again use the home agent to supply a registration key.

4. If the mobile node includes its public key in its Registration Request, the foreign agent can choose the new registration key.

5. The mobile node and its foreign agent can execute a Diffie–Hellman key exchange protocol [5] as part of the registration protocol.

If a request for key establishment cannot be accommodated by the foreign agent and/or the home agent, then the mobile node's key request must go unfulfilled. This does not mean that the Registration Request itself fails, so it has no effect on the status code returned by the home agent to the mobile node. The mobile node must be able to handle the case in which it has requested a key but the Registration Reply arrives without any key reply extension. This could happen even when the foreign agent has advertised its willingness to offer smooth handoffs, and the mobile node has supplied all the necessary parameters (e.g., for a Diffie–Hellman key exchange).

Once the registration key is established, the method for performing smooth handoff is natural, as described earlier. In the remainder of this section, we give a brief overview of the proposed methods for establishing the registration key enumerated above.

6.1. The home agent as a KDC

The second and third methods listed above make use of the mobility security association shared between the home agent and mobile node, to encode the registration key for delivery to the mobile node. Thus, if the home agent can securely deliver the key to the foreign agent, it can be used as a *Key Distribution Center* (KDC) for the mobile node and its new foreign agent. The mobile node requests this by including a *Registration Key Request* extension in its Registration Request message. When the home agent chooses the registration key, it returns the key in two different extensions to the Registration Reply. One extension has the key encrypted for the foreign agent, and the other extension has the same key encrypted differently for the mobile node.

For the registration key to be established using this method, the home agent must be able to securely transmit an encrypted copy of the registration key to the foreign agent. This is straightforward if the foreign agent already has a mobility security association with the home agent. If mobile nodes from some home network often visit a foreign agent, then the effort of creating such a mobility security association between that foreign agent and the home agent serving their home network may be worthwhile.

If no such mobility security association exists, but the foreign agent has a public key available, it can still ask the home agent to use it to pick a registration key. This may be preferable to asking the mobile node to pick a good registration key, because doing so may depend upon

RED HAT 1010
RED HAT V. CAS

using resources not available to all mobile nodes; simply selecting cryptographically strong pseudo-random numbers may by itself be a significant computational burden [6]. Moreover, allowing the home agent to pick the key fits well into the existing registration procedures. On the other hand, it is possible that a mobile node could do with less than perfect pseudo-random numbers as long as the registration key were to be used in the restricted fashion envisioned here for smooth handoffs.

6.2. Using the foreign agent as a KDC

When the foreign agent and mobile node share a mobility security association, there is no need to pick a registration key. The mobile node can secure its Binding Update to the foreign agent whenever it needs to, by using the existing security association. This is the most desirable case.

Otherwise, if available, the mobile node can include its public key (such as RSA [18]) in its Registration Request to the foreign agent, using a *Mobile Node Public Key* extension. The foreign agent chooses the new registration key and includes a copy of it in the Registration Request, encrypted with the mobile node's public key, using a *Foreign-Mobile Registration Key Reply* extension.

6.3. Using Diffie–Hellman with the foreign agent

The Diffie–Hellman key-exchange algorithm [5,18] can be used. Diffie–Hellman is a public key cryptosystem that allows two parties to establish a shared secret key, such that the shared secret key cannot be determined by other parties overhearing the messages exchanged during the algorithm. It is already used, for example, in other protocols that require a key exchange, such as in the *Cellular Digital Packet Data* (CDPD) system [2].

Some applications of this technique are known to suffer from a *man-in-the-middle* attack. In other words, a malicious agent *could* pretend to the foreign agent to be the mobile node, and pretend to the mobile node to be the foreign agent, and participate as an unwanted third member in the key exchange. Armed with knowledge of the registration key, the malicious agent *could* at a later time disrupt the smooth handoff, or initiate the handoff prematurely. In Route Optimization, Diffie–Hellman results are authenticated by the home agent to frustrate such man-in-the-middle attacks. Moreover, the mobile node and/or the foreign agent are presumably in direct contact, so that such an attack is detectable if either of the nodes notices the reception of duplicate packets, and corrective action taken.

If Diffie–Hellman were not computationally expensive, it could likely serve the needs of most mobile nodes. But, the algorithm itself uses exponentiations involving numbers with hundreds of digits. That may take a long time for some mobile nodes to compute, time which might come at the expense of interactivity or convenient operation of user application programs. For this reason, Diffie–Hellman may be less desirable than some other methods for establishing registration keys. Since it requires no other configuration, it is nevertheless required in all implementations of foreign agents that advertise support for smooth handoffs.

Briefly, the Diffie–Hellman algorithm involves the use of two large public numbers, a *prime number* and a *generator*. The prime number and the generator must be known by both parties involved in the algorithm, but do not have to be secret; these values may be the same or different for each execution of the algorithm and are not used once the algorithm completes. Each party chooses a private random number, produces a *computed value* based on this random number, the prime and the generator, and sends the computed value in a message to the other party. Each party then computes the (same) shared secret key using its own private random number, the computed value received from the other party, and the prime and generator values. Knowing the computed values and not the chosen random numbers does not enable passive listeners to determine the shared secret key.

To use the Diffie–Hellman algorithm during registration with a foreign agent, the mobile node includes a *Registration Key Request* extension in its Registration Request message, containing its nonzero values for the prime and generator, along with the computed value from its own private random number. The foreign agent then chooses its own private random number and includes a *Diffie–Hellman Registration Key Reply* extension in its Registration Reply message to the mobile node; the extension includes the foreign agent's own computed value based on its chosen random number and the supplied prime and generator values from the mobile node. The mobile node and foreign agent each independently form the (same) shared secret key from their own chosen random number, the computed value supplied by the other party, and the prime and generator values.

Establishing a registration key using Diffie–Hellman is computationally expensive, but the use of Diffie–Hellman described here is designed to allow the Diffie–Hellman computations to be overlapped with other activities. The mobile node may choose (or be manually configured with) the prime and generator values at any time, or may use the same two values for a number of registrations. The mobile node may also choose its private random number and calculate its computed value at any time. For example, after completing one registration, the mobile node may choose the private random number for its next registration and begin the computation of its new computed value based on this random number, such that it has completed this computation before it is needed in its next registration. Even more simply, the mobile node may use the same private random number and computed value for any number of registrations. The foreign agent may choose its private random number and begin computation of its computed value based on this number as soon as it receives the mobile node's Registration Request message.

RED HAT 1010
RED HAT V. CAS

This method could be extended to support other similar key exchange algorithms, either by adding a new request and reply extension for each, or by adding a field in the extensions to indicate which algorithm is to be used.

7. Messages requesting a registration key

This section describes the extensions that may be used by mobile nodes or foreign agents to request the establishment of a registration key. See section 9 for appropriate algorithms which allow each node to tailor the use of these extensions to most closely fit its configured requirements.

7.1. Foreign Agent Key Request extension

If the foreign agent receives a Registration Key Request from a mobile node and it has a security association with the home agent, it may append the *Foreign Agent Key Request* extension to the Registration Request after the *Mobile-Home Authentication* extension. The home agent will use the SPI specified in the key request extension to encode the registration key in the subsequent Registration Reply message. The format of the Foreign Agent Key Request extension is illustrated in figure 11.

7.2. Mobile Node Public Key extension

If the mobile node has a public key, it can ask its prospective foreign agent to choose a registration key, and to use the mobile node's public key to encode the chosen registration key. No eavesdropper will be able to decode the registration key, even if it is broadcast to all entities with access to the network medium used by the mobile node. If using the public key, the foreign agent should still include the selected key in the Registration Request before it goes to the home agent. Then, the home agent can authenticate the selected encoded registration key as part of the Registration Reply message. The format of the *Mobile Node Public Key* extension is illustrated in figure 12.

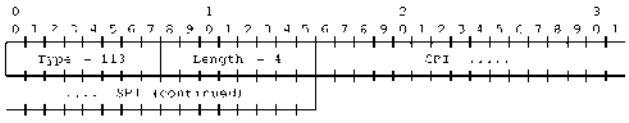


Figure 11. Foreign Agent Key Request extension format.

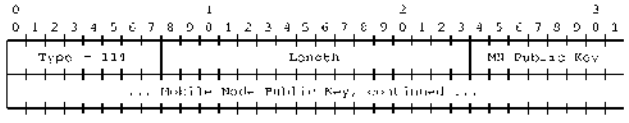


Figure 12. Mobile Node Public Key extension format.

7.3. Foreign Agent Public Key extension

The format of the *Foreign Agent Public Key* extension is illustrated in figure 13. Using this extension, if the foreign agent has a public key, it can ask the home agent to choose a registration key, and to use the foreign agent's public key to encode the chosen registration key. Then, the home agent can authenticate the selected encoded registration key as part of the Registration Reply message. The SPI field in the extension is provided for the home agent to transcribe into the eventual *Foreign Agent Public Key Reply* extension to the Registration Reply message.

7.4. Registration Key Request extension

The *Registration Key Request* extension, illustrated in figure 14, may be included in a Registration Request message sent to a foreign agent. If the length of the parameters in the key request extension are all zero, then the mobile node is asking the foreign agent to supply a key by any means it has available except for Diffie–Hellman.

If the lengths are nonzero, then the mobile node is enabling the foreign agent to also perform the Diffie–Hellman key exchange algorithm (as described in section 6.3) if the other possible key establishment methods are not available. The foreign agent should then select a good pseudo-random registration key, and include a *Diffie–Hellman Registration Key Reply* extension, in the Registration Request message sent to the home agent to complete the key exchange. The home agent will also include the same extension in the Registration Reply sent to the mobile node, which will then be authenticated as part of the reply message.

The *Prime* and *Generator* fields give the two public values for this execution of the Diffie–Hellman algorithm. The *Computed Value* is the public computed value from the mobile node for this Diffie–Hellman exchange. The values *Prime*, *Generator*, and *Computed Value* must all be the same length, which must be a multiple of 8 bits. Correspondingly, the *Length* field is a number which is 3 times the length (in bytes) of each of *Prime*, *Generator*, and *Computed Value*.

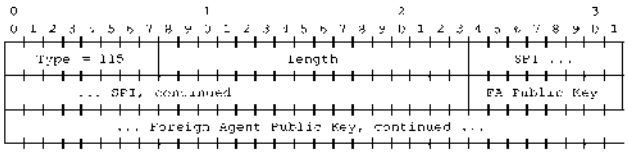


Figure 13. Foreign Agent Public Key extension format.

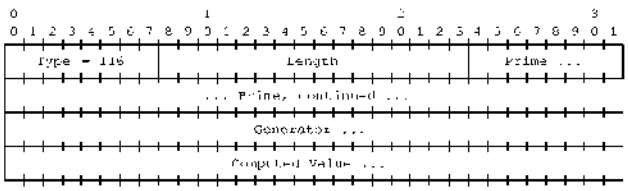


Figure 14. Registration Key Request extension format.

RED HAT 1010
RED HAT V. CAS

8. Extensions to supply a registration key

This section describes the extensions that may be used to supply a registration key to a requesting entity, either a foreign agent or a mobile node. These extensions are the counterparts to the corresponding extensions used to request registration keys that were described in the last section.

8.1. Home-Mobile Key Reply extension

The *Home-Mobile Key Reply* extension may be used in Registration Reply messages to send a registration key from the mobile node's home agent to the mobile node. When used, the home agent is required to also include a key reply extension in the Registration Reply message, which gives a copy of the same key to the mobile node's new foreign agent. The Home-Mobile Key Reply extension, illustrated in figure 15, is authenticated along with the rest of the Registration Reply message, and thus no additional authenticator is included in the extension. The SPI used to encode the registration key may be different from the SPI used to authenticate the Registration Reply message.

The *Mobile Node Encrypted Key* is the registration key, chosen by the home agent, encrypted under the mobility security association between the home agent and the mobile node. The same key must be sent, encrypted for the foreign agent in a Foreign Agent Key Reply extension in this Registration Reply message.

8.2. Foreign Agent Key Reply extension

The *Foreign Agent Key Reply* extension, illustrated in figure 16, may be used in Registration Reply messages to send a registration key from the mobile node's home agent to the mobile node's new foreign agent. The registration key is encrypted under the mobility security association between the home agent and the foreign agent. The key that is sent in this extension must also be sent by the home agent to the mobile node, encoded for the mobile node in a *Home-Mobile Key Reply* extension in the same Registration Reply message.

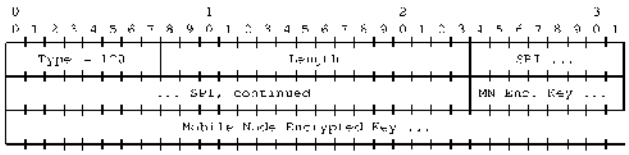


Figure 15. Home-Mobile Key Reply extension format.

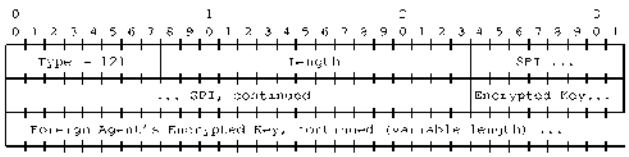


Figure 16. Foreign Agent Key Reply extension format.

Authentication of the key is performed by use of data within a Home-Foreign Authentication extension to the Registration Reply message, which is required when the Foreign Agent Key Reply extension is used. Replay protection is accomplished using the Identification field in the Registration Request message, which is also used by the foreign agent to identify the pending registration data.

8.3. Mobile Node Public Key Reply extension

The Mobile Node Public Key Reply extension is illustrated in figure 17. When the mobile node sends a *Mobile Node Public Key Request* to its prospective foreign agent, the foreign agent can immediately select a registration key. The foreign agent encodes this registration key into the *Mobile Node Public Key Reply* extension to the Registration Request, along with an SPI for future reference to the key. The home agent subsequently transcribes the extension without change into the Registration Reply message.

8.4. Foreign Agent Public Key Reply extension

In response to a *Foreign Agent Public Key Request* extension, the home agent will select a registration key and encode it into two separate key reply extensions of the Registration Reply message. The Foreign Agent Public Key Reply message is illustrated in figure 18. The Foreign Agent Public Key Reply extension contains the registration key encrypted with the public key of the foreign agent. The *Foreign Agent's Encrypted Key* is a pseudo-random number, chosen by the home agent, and encrypted using the foreign agent's public key. The SPI, provided by the foreign agent for transcribing into this extension, is ultimately targeted for use by the mobile node.

8.5. Diffie–Hellman Key Reply extension

The *Diffie–Hellman Registration Key Reply* extension, illustrated in figure 19, should be included in a Registration Request message sent by a foreign agent to the home agent, when the following conditions are met:

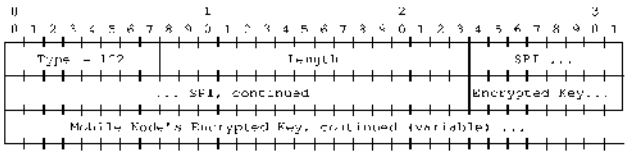


Figure 17. Mobile Node Public Key Reply extension format.

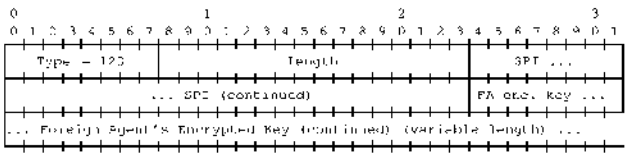


Figure 18. Foreign Agent Public Key Reply extension format.

RED HAT 1010
RED HAT V. CAS

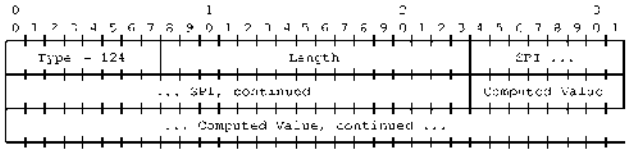


Figure 19. Diffie–Hellman Key Reply extension format.

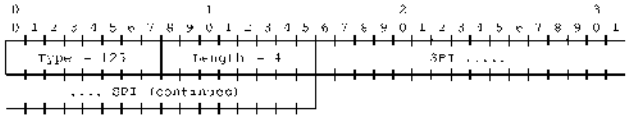


Figure 20. SPI extension format.

- the mobile node has included a *Registration Key Request* extension with nonzero prime and generator in its Registration Request message to the foreign agent, and
- the foreign agent has no public key or security association with the home agent or mobile node.

The *Computed Value* field gives the computed value from the foreign agent for this execution of the Diffie–Hellman algorithm. The values of the prime and generator are taken from the Registration Key Request extension from the mobile node's Registration Request message. The foreign agent supplies a new SPI along with the new registration key, so that the new key will be useful in the same way as registration keys created by any other method.

8.6. SPI extension

The SPI extension is included in Registration Reply messages when needed to specify the SPI to be associated with the registration key, which may be different from the SPI used to encrypt the key. The format of the SPI extension is shown in figure 20.

9. Registration Key Request processing

This section describes the processing steps taken by a mobile node, foreign agent, and home agent in establishing a registration key for the mobile node during registration. These steps use the message extensions described in section 8.

9.1. Mobile Node Key Requests

If the mobile node receives an Agent Advertisement from a foreign agent with the *smooth handoff* bit set, it may initiate a smooth handoff with its previous foreign agent, as well as asking its new foreign agent to aid in supplying a registration key for the new registration. The following algorithm may be used by the mobile node during registration to allow flexibility in the selection of the new registration key. Any particular mobile node may be configured to use one, none, or any subset of the key establishment procedures made available as part of the Route Optimization protocol.

1. If the *smooth handoff* bit is not set in the Agent Advertisement, do not include a Registration Key Request in the Mobile IP registration message, and skip all of the following steps.
2. If the mobile node has a registration key with the previous foreign agent, append the Previous Foreign Agent Notification extension.
3. If a security association exists with the new foreign agent, use it to secure any future Previous Foreign Agent Notification extensions.
4. Otherwise, if a public key is owned by the mobile node, append the Mobile Node Public Key Request extension.
5. Otherwise, if a Diffie–Hellman key exchange is desired, include a value for the prime, generator, and mobile node computed value in a Mobile Node Key Request extension.
6. Otherwise, append the Registration Key Request extension with empty (zero-length) prime, generator, and computed value.

In this way, the mobile node can get a registration key whenever one can be produced by any of the several available mechanisms.

9.2. Home agent processing for Registration Key Requests

When the home agent receives a Registration Request message, an extension requesting a registration key may be present in the message, asking the home agent to provide a registration key to the mobile node and its foreign agent, as described in section 6. In that event, the home agent employs a good algorithm for producing random keys [6] and encrypts the result separately for use by the foreign agent and by the mobile node. The chosen key is encrypted under the mobility security association shared between the home agent and the mobile node, and the encrypted key is placed in a Home-Mobile Registration Key Reply extension (section 8.1) in the Registration Reply message. The same key is also encrypted under the mobility security association shared between the home agent and the foreign agent, and the encrypted key is placed in a Home-Foreign Registration Key Reply extension (section 8.2) in the Registration Reply message.

If the home agent cannot satisfy a request to select a registration key, it may still satisfy the registration attempt. In this case, the home agent returns a Registration Reply message indicating success, but does not include any key reply extension.

9.3. Home agent supplying registration keys

When the home agent receives a Registration Request message with registration key extensions, it usually performs one of two operations:

RED HAT 1010
RED HAT V. CAS

- the home agent selects and encodes a registration key for both the mobile node and the foreign agent, or
- it transcribes the registration key already selected by the foreign agent into the appropriate extension to the Registration Reply message.

Either operation ensures that the mobile node and home agent are dealing with the same foreign agent.

When building the Registration Reply, the home agent should follow an algorithm such as the one suggested below, to be useful for the range of registration key establishment scenarios that are possible given the current route optimization protocol.

1. If the Registration Request contains extensions indicating that the foreign agent has already selected a registration key for use by the mobile node, copy those extensions into the Registration Reply.
2. Otherwise, if the Registration Request contains a Foreign Agent Key Request, then there is a security association with the foreign agent; append the Mobile Node Key Reply extension and the Foreign Agent Key Reply extension to the Registration Reply.
3. Otherwise, if a public key is available for the foreign agent (perhaps contained in the Foreign Agent Public Key extension), append the Mobile Node Key Reply extension and the Foreign Agent Public Key Reply extension to the Registration Reply.
4. Otherwise, do nothing. A registration key cannot be provided, but the registration can still be accepted according to the base Mobile IP specification.

In all cases, the home agent authenticates the Registration Key Reply extensions using the Home-Mobile Authentication extension.

9.4. Foreign agent processing for key requests

The foreign agent, when it receives a request from a mobile node for a registration key, is faced with a variety of possible actions. The action selected by the foreign agent depends on the resources it has available. The foreign agent typically attempts to reduce as much as possible the computational burden placed on the mobile node, but relies on the security association with the greatest cryptographic strength to encode the registration key. Furthermore, if the foreign agent performs the key selection, it still supplies the encoded key in an extension to the Registration Request message, so that the process of registration will also have the effect of authenticating its choice of registration key to the mobile node. This strategy reduces the opportunity for attackers to mount *man-in-the-middle* attacks.

The following steps illustrate an algorithm that may be used when a foreign agent gets a Registration Request with one of the key request extensions included.

1. If the Registration Request contains a Previous Foreign Agent Notification extension, perform the indicated function by building and transmitting the appropriate Binding Update message to the previous foreign agent.
2. If there is a security association with the home agent, append a Foreign Agent Key Request extension to the Registration Request.
3. Otherwise, if the foreign agent has a public key, append the Foreign Agent Public Key Request to the the Registration Request.
4. Otherwise, if the mobile node's public key is available, use it to pick a good registration key and append the Foreign Agent Public Key Reply extension.
5. Otherwise, if the mobile node has indicated that a Diffie–Hellman key exchange is desired, perform the computation and use the result when appending the Diffie–Hellman Key Reply extension to the Registration Request.
6. Otherwise, forward the Registration Request to the home agent without any further changes.

10. Route Optimization in IPv6

The development of Route Optimization techniques for IPv4 has played an important role in the development of mobility support for IP Version 6 (IPv6) [4,10]. Most IPv4 correspondent nodes will not support the processing of Binding Updates for years. IPv6 correspondent nodes, on the other hand, should be able to support the analogous features in IPv6 much sooner. There are three reasons for the difference:

- There is not a sizable deployment of IPv6 nodes to impede the introduction of appropriate protocol features.
- Efficient support of mobility was mandated by the IPng Directorate and the IESG (Internet Engineering Steering Group) [1].
- The base IPv6 protocol supports mobility more naturally than IPv4, so that there is less work needed overall to create products that implement the standard.

One of the main features of IPv6 that is most useful for mobility support is the requirement that all IPv6 nodes support authentication. Since the nodes have to perform authentication, APIs have been developed to enable the functions at the network protocol layer, and the APIs can be used to do key management. With keys distributed to the correspondent nodes, the mobile node can expect its correspondents to accept authenticated Binding Updates. When the correspondents are able to keep track of the mobile node's care-of addresses, they can (just as in IPv4 Route Optimization) cause packets to go directly to the mobile node without ever passing through the home network or home agent.

RED HAT 1010
RED HAT V. CAS

There are further advantages inherent in IPv6 for mobility support, many of which are not relevant to Route Optimization. One important feature, however, is the new organization for IPv6 options. Options can now be specified to be end-to-end, instead of hop-by-hop as with IPv4. For this reason, in IPv6, Binding Updates are delivered in end-to-end options (called *destination options*). Moreover, since normal packets can carry destination options, Binding Updates can be delivered to correspondent nodes without adding additional control packets to the network, thus reducing network load. Moreover, since intermediate nodes (routers) do not process the destination option, there is little routing penalty for the transmission of binding updates to correspondent nodes.

Given the expected universal deployment of authentication algorithms in IPv6 nodes, it is more reasonable to expect that a mobile node can maintain a security association with all its correspondents. For this reason, Binding Updates are delivered directly from the mobile node to the correspondent node in IPv6, in contrast to the method used in IPv4 whereby home agents are responsible for this chore. The mobile node can tell directly when a correspondent node needs a Binding Update. Whenever the mobile node moves to a new point of attachment to the Internet, any of its correspondents with active connections will need to get the new care-of address. The mobile node then inserts a Binding Update destination option into the next packet that it needs to transmit to any particular correspondent. Furthermore, the mobile node can tell if an incoming packet was originated by a correspondent node with an out-of-date or nonexistent binding for the mobile node, since such packets have to be tunneled to the mobile node from the home agent instead of directly from the correspondent node.

Lastly, the IPv6 analog for smooth handoffs from one point of attachment to the next deserves mention. Since each mobile node can acquire a care-of address by using Neighbor Discovery [12] and IPv6 address autoconfiguration [19], there is no need for any foreign agent to supply this information to the mobile node. With 128-bit addresses, there is not the same incentive for conserving the IPv6 address space, either. So, when a mobile node moves to a new network, there is no foreign agent left behind to help with forwarding data packets still in flight to the old network.

To solve this problem, the mobile node can deliver a Binding Update for its previous care-of address to the default router on its previous network. This default router need not even know anything about the mobile node's home address. In fact, the previous router acts as a home agent for the mobile node's previous care-of address.

11. Conclusions

In this paper, we have presented the current proposed protocol definition for *Route Optimization*, by which is meant the elimination of *triangle routing* whenever the correspondent node is able to perform the necessary protocol operations. The Route Optimization protocol definition is largely concerned with supplying a *Binding Update* to any correspondent node that needs one (and can process it correctly). The Binding Update message is also used in conjunction with the Previous Foreign Agent Notification extension to allow for smooth handoffs between foreign agents.

Furthermore, we have presented some methods for establishing registration keys for use by mobile nodes and foreign agents supporting smooth handoffs. We have given detailed processing requirements for mobile nodes, foreign agents, and home agents. Finally, the essential features of Route Optimization, as realized in IPv6, have been identified. We have discussed the differences between IPv4 and IPv6, with the hope that in so doing, the design space for Route Optimization will be more fully understood, and that IPv6's ability to support mobility will be more fully appreciated.

Acknowledgements

The work of David Johnson in this paper was supported in part by the National Science Foundation (NSF) under CAREER Award NCR-9502725, by the Air Force Materiel Command (AFMC) under DARPA contract numbers F19628-93-C-0193 and F19628-96-C-0061, and by the AT&T Foundation under a Special Purpose Grant in Science and Engineering. The views and conclusions contained here are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either express or implied, of NSF, AFMC, DARPA, the AT&T Foundation, Sun Microsystems, Carnegie Mellon University, or the U.S. Government.

References

[1] S. Bradner and A. Mankin, The recommendation for the IP Next Generation Protocol, RFC 1752 (January 1995).

[2] CDPD consortium, *Cellular Digital Packet Data Specification* (Chicago, IL, 1993).

[3] S.E. Deering, ed., ICMP router discovery messages, RFC 1256 (September 1991).

[4] S. Deering and R. Hinden, Internet Protocol, version 6 (IPv6) specification, RFC 1883 (December 1995).

[5] W. Diffie and M. Hellman, New directions in cryptography, IEEE Transactions on Information Theory 22 (November 1976) 644–654.

[6] D.E. Eastlake, S.D. Crocker and J.I. Schiller, Randomness recommendations for security, RFC 1750 (December 1994).

[7] S. Hanks, T. Li, D. Farinacci and P. Traina, Generic Routing Encapsulation (GRE), RFC 1701 (October 1994).

[8] S. Hanks, T. Li, D. Farinacci and P. Traina, Generic routing encapsulation over IPv4 networks, RFC 1702 (October 1994).

[9] D.B. Johnson, Scalable and robust internetwork routing for mobile hosts, in: *Proceedings of the 14th International Conference on Distributed Computing Systems* (June 1994) pp. 2–11.

[10] D. Johnson and C. Perkins, Mobility support in IPv6, draft-ietf-mobileip-ipv6-06.txt (August 1998). (Work in progress.)

[11] H. Krawczyk, M. Bellare and R. Canetti, HMAC: keyed-hashing for message authentication, RFC 2104 (February 1997).

RED HAT 1010
RED HAT V. CAS

[12] T. Narten, E. Nordmark and W. Simpson, Neighbor discovery for IP version 6 (IPv6), RFC 1970 (August 1996).
[13] C. Perkins, IP encapsulation within IP, RFC 2003 (May 1996).
[14] C. Perkins, Minimal encapsulation within IP, RFC 2004 (May 1996).
[15] C. Perkins, ed., IP mobility support, RFC 2002 (October 1996).
[16] C.E. Perkins and D.B. Johnson, Special tunnels for Mobile IP, draft-ietf-mobileip-specun-00.txt (November 1997). (Work in progress.)
[17] J.B. Postel, ed., Internet control message protocol, RFC 792 (September 1981).
[18] B. Schneier, *Applied Cryptography: Protocols, Algorithms, and Source Code in C* (Wiley, New York, 2nd ed., 1996).
[19] S. Thomson and T. Narten, IPv6 stateless address autoconfiguration, RFC 1971 (August 1996).

Charles E. Perkins is a Senior Staff Engineer at Sun Microsystems, developing Service Location Protocol and investigating dynamic configuration protocols for mobile networking. He is the editor for ACM/IEEE Transactions on Networking and for ACM/URSI/Baltzer Wireless Networks in the area of wireless networking. He is serving as document editor for the mobile-IP working group of the Internet Engineering Task Force (IETF), and is author or co-author of standards-track documents in the mobileip, svrloc, dhc (Dynamic Host Configuration) and IPng working groups. Charles is also associate editor for Mobile Communications and Computing Review, the official publication of ACM SIGMOBILE. He is serving on the Internet Architecture Board (IAB). Charles has authored a book on Mobile IP, and has published a number of papers in the areas of mobile networking, ad-hoc networking, route optimization for mobile networking, resource discovery, and automatic configuration for mobile computers. He is lead guest editor of an upcoming issue of the ACM/Baltzer journal Mobile Networks and Applications. Charles has served on various committees for the National Research Council, and is currently the chairperson of the Nomadicity Working Team of the Cross-Industry Working Team (XIWT). His previous projects included developing multiprocessor operating systems using Mach, and adapting Unix to fit on PDAs (personal digital assistants), multiprocessor systems and user interface prototyping systems. Charles holds a B.A. in mathematics and a M.E.E. degree from Rice University, and a M.A. in mathematics from Columbia University. He is a member of ISOC, ACM, IEEE, and the IETF.
E-mail: cperkins@eng.sun.com

David B. Johnson is an Assistant Professor in the School of Computer Science at Carnegie Mellon University. He also holds a courtesy faculty appointment in the Electrical and Computer Engineering Department at Carnegie Mellon, and is a member of CMU's Information Networking Institute. His research interests include network protocols, distributed systems, and operating systems. Prior to joining the faculty at CMU in 1992, he was on the faculty at Rice University for three years as a Research Scientist and Lecturer in the Computer Science Department. He received a B.A. in computer science and mathematical sciences in 1982, an M.S. in computer science in 1985, and a Ph.D. in computer science in 1990, all from Rice University. Professor Johnson is currently leading the Monarch Project at CMU, developing adaptive networking protocols and protocol interfaces to allow truly seamless wireless and mobile host networking. Related to this research, he has been active in the Internet Engineering Task Force (IETF) for many years and is one of the principal designers of the IETF Mobile IP protocol for IPv4 and IPv6. In 1997, he was Program Chair for the Third Annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom '97) and is currently Technical Vice Chair for Mobile Systems for the 19th International Conference on Distributed Computing Systems (ICDCS '99). He is Executive Committee member and the Treasurer for the ACM Special Interest Group on Mobile Computing and Communications (SIGMOBILE), an Area Editor for the ACM/Baltzer journal Mobile Networks and Applications (MONET) and the ACM SIGMOBILE magazine Mobile Computing and Communications Review (MC2R), and a Guest Editor for an upcoming issue of IEEE Journal on Selected Areas in Communications (J-SAC) on mobile ad hoc networking. He is a member of the IEEE Computer Society, IEEE Communications Society, ACM, USENIX, Sigma Xi, and the Internet Society.
E-mail: dbj@cs.cmu.edu

RED HAT 1010
RED HAT V. CAS

Network Working Group C. Huitema

Internet Draft INRIA

Expiration Date: November 1995 May 1995

Multi-homed TCP

C. Huitema

Christian.Huitema@sophia.inria.fr

draft-huitema-multi-homed-0.txt

1. Status of this Memo

This document is an Internet-Draft. Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet- Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet- Drafts as reference material or to cite them other than as ''work in progress.''

To learn the current status of any Internet-Draft, please check the ''lid-abstracts.txt'' listing contained in the Internet- Drafts Shadow Directories on ftp.is.co.za (Africa), nic.nordu.net (Europe), munnari.oz.au (Pacific Rim), ds.internic.net (US East Coast), or ftp.isi.edu (US West Coast).

RED HAT 1010
RED HAT V. CAS

2. Abstract

Current TCP connections are identified by pairs of host addresses and port numbers. This introduces a "fate sharing" dependency between the connection and these values, and specially with the time to live of the host addresses. There are at least three cases where this dependency is unduly harmful:

(1) when the host moves to a new location and is assigned a new address,

(2) when the interface used by a multi-homed host to initiate the connection is temporarily disconnected,

(3) when the host's network is renumbered, e.g. after changing provider.

We propose to remove this fate sharing effect by allowing the set of addresses used by a TCP connection to change over time. We modify TCP defining a new type of parameter, PCB-ID, to be used during the initial synchronisation. This parameter identifies the "Protocol Control Block" associated to the TCP connection. When initiating a connection, the host attach to the SYN packet the identifier of the local PCB. If both hosts have identified their local PCB, they can now exchange "Extented TCP" packets, where the pair of 16-bit port numbers has been replaced by the

RED HAT 1010
RED HAT V. CAS

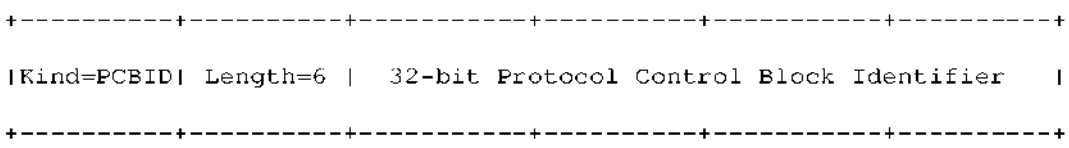
32-bit PCB-ID at the destination. This way, PCB location becomes independent of the IP addresses. The addresses in use can be stored in this PCB, so that we can change these in the course of the connection. In fact, we can use several addresses in parallel for the same connection, which is why our proposal is called "multi-homed TCP".

3. Proposed extensions to TCP

Our proposal requires two extensions to TCP: the definition of a context parameter during the synchronization exchange and the carrying of this parameter in the data packets.

3.1. The context identification parameter

When a TCP entity initiates a new connection, it can announce its willingness to receive data from alternate addresses by including a "PCB identification" parameter in the initial "synchronization" packet. This parameter has the following format:



This option is a unilateral offer, not subject to negotiation. It allows the partner to send data from an alternate source address, if indeed the "local contex id" is somehow repeated in the new packets. The context-id is normally the identification of a "protocol control block"; however, the precise format and bits carried in this parameter are entirely a local matter. The only condition is one on reuse: in order to avoid protocol errors due to old packets popping out of the network, the context-id used by one connection cannot be reused immediately. One must

RED HAT 1010
RED HAT V. CAS

wait for the end of the "TIME- WAIT" state.

The parameter kind PCBID is to be allocated by the IANA.

This parameter can only be present in TCP packets where the SYN bit is set.

3.2. Incoming data

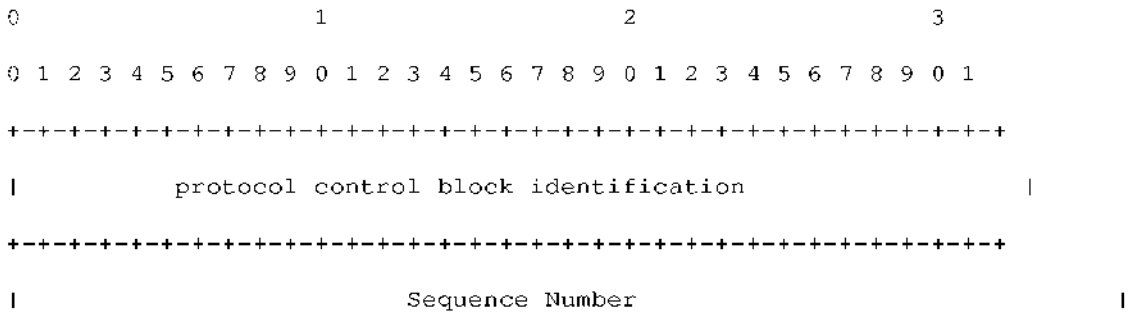
An entity that has learned the PCB identification of its partner can elect to send "Extended TCP" packets. The difference between extended TCP and regular TCP packets are that:

- the IP protocol type is (a value to be defined by IANA), not 7 (TCP).

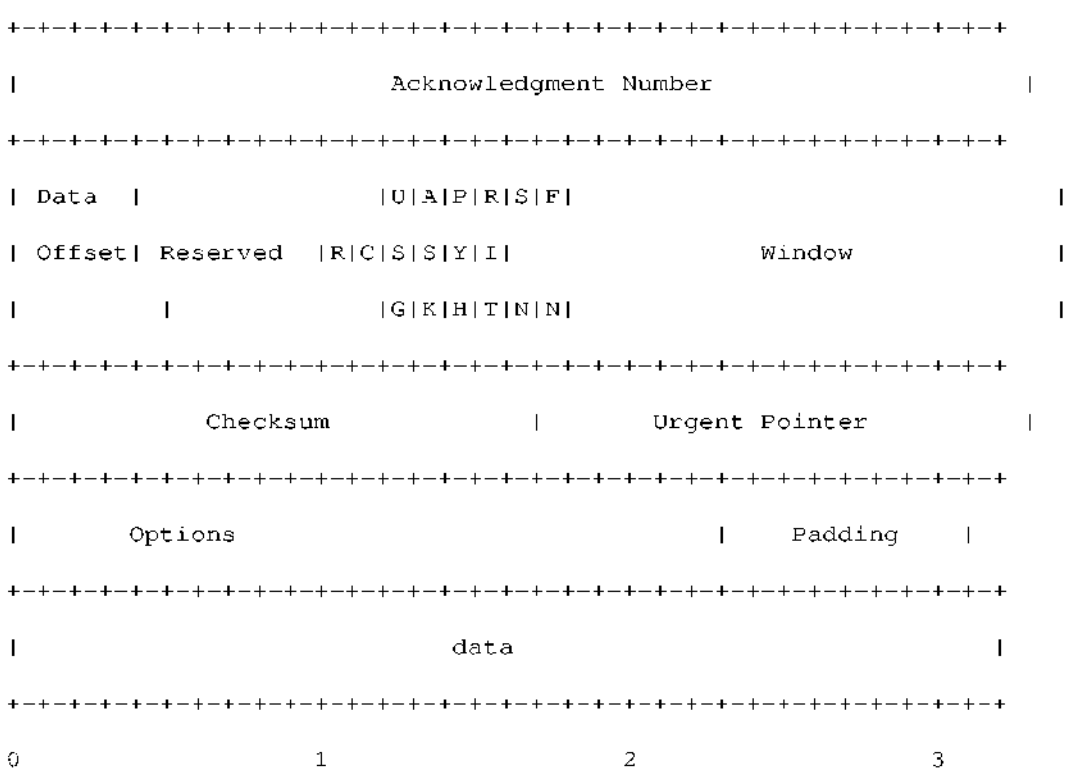
- the IP source and destination addresses may be different from the ones use in the SYN packet.

- the first 32 bits of the ETCP header carry a protocol control block identification, not a pair of ports.

The extended TCP packets have a format identical to normal TCP packets, except for the replacement of the port pair by a context identification:



RED HAT 1010
RED HAT V. CAS



ETCP Header Format

ETCP packets are delivered to an ETCP incoming process, which merely looks at the context identification and finds the corresponding PCB. All packets arriving on the PCB identified by the context XB are treated exactly as if they had been sent by host A to host B, using the TCP ports "a" and "b", i.e. the ports present in the TCP packet.

3.3. Extended checksum rules

The checksum of ETCP packets is computed over the pseudo-header and the packet itself. The pseudo header is composed of the IP addresses actually present in the packet.

3.4. PCB extension

RED HAT 1010
RED HAT V. CAS

The protocol control block of TCB is extended to include a list of partners addresses. This list is initialized with the partner address expressed in the connection request. It is populated with the source addresses from which ETCP packets bound to that PCB are received. Two dates are associated to each address, "first seen" and "last seen", i.e. the date at which a packet was received from this address the first time, and then the last time.

The list is sorted by reverse order of "last seen" packets, i.e. most recently seen first. TCP entities should normally send their packets to the most recently seen address. They may elect to spread them over all of the "currently active" addresses, i.e. all addresses whose "last seen" date is larger than the "first seen" date of the most recently seen address.

4. Finding addresses

The TCP "protocol control block" will be extended to contain a list of IP addresses of the source entity. These addresses are obtained by monitoring the IP source addresses of the ETCP packets.

In this section, we will see the address reassignent at use in four cases, deprecation, anycasting, multi-homing and mobility.

4.1. Deprecation

The address autoconfiguration and neighbor discovery procedures of IPv6 allows hosts to renumber their interface. This is supposed to occur in three phases:

RED HAT 1010
RED HAT V. CAS

1) Initially, the interface of host A is numbered with the address A1,

2) After reception of a router advertisement message, the host A decide that the preferred address of the interface is now A2. Packets sent to the deprecated address A1 can still be received for some period.

3) At the end of that period, the address A1 is invalidated. Packets sent to it are discarded by the network.

Let's discuss first the behaviour of the host A, which is engaged in a TCP connection with the host B, using ports a and b and PCB identifiers xa and xb, respectively. Initially, packets will be exchanged using the addresses A1 and B1.

```
(A1, B1, tcp) (a, b, SYN, PCBID=xa) ===========>
      <============ (B1, A1, etcp) (xa, SYN, ACK, PCBID=xb)
(A1, B1, etcp) (xb, ACK, DATA) ===========>
      <============ (B1, A1, etcp) (xa, ACK)
```

As soon as the address A1 becomes deprecated, A will stop using it. The next data or acknowledgement packets will be sent from the preferred source address A2:

```
(A2, B1, etcp) (xb, DATA) ===========>
      <============ (B1, A1, etcp) (xa, DATA)
(A2, B1, etcp) (xb, ACK, DATA) ===========>
      <============ (B1, A2, etcp) (xa, ACK)
```

The address change is noted by B because upon reception in context "xb" of a packet from the new address A2. At this stage, B switches to the new address. Due to the propagation delay, some packets sent to the old

RED HAT 1010
RED HAT V. CAS

address may still be in transit at this point. This will last for at most twice the maximum propagation delay. We may then ensure that no packets will be lost if the old address remains valid for at least twice that delay, e.g. more than four minutes.

4.2. Anycasting

In order to use IPv6's anycasting facility, we should be able to send the initial SYN packet to an anycast address. Such connection requests are doomed in today's TCP, because anycast addresses cannot be used as source addresses. ETCP solves the problem rather elegantly. Suppose that the packet sent by the host A to the anycast address X is actually routed to the nearest server for that anycast address, B, through its interface B1.

```
(A1, X, tcp) (a, b, SYN, PCBID=xa) ===========>
       <============ (B1, A1, etcp) (xa, SYN, ACK, PCBID=xb)
(A1, B1, etcp) (xb, ACK, DATA) ===========>
       <============ (B1, A1, etcp) (xa, ACK)
```

Because B knows the PCB identifier "xa", it can sent an ETCP packet back towards A, using the regular address B1 as source address, advertizing its own PCB identifier. A notes the source address B1 in its PCB, and the following packets are happily exchanged between A1 and B1.

The solution can in fact be extended to other variations of anycasting, e.g. when the initial packet is sent to a multicast group. The only difference here is that several response will come, e.g. from B and C, two members of the group which both decided to reply to A.

RED HAT 1010
RED HAT V. CAS

```
          (A1, G, tcp) (a, b, SYN, PCBID=xa) ===========>
                 <============ (B1, A1, etcp) (xa, SYN, ACK, PCBID=xb)
          (A1, B1, etcp) (xb, ACK, DATA) ===========>
                      <============ (C1, A1, etcp) (xa, SYN, ACK,
PCBID=xc)
          (A1, C1, etcp) (xc, RST) ===========>
          (A1, B1, etcp) (xb, DATA) ===========>
```

Upon reception of the synchronisation packet from C, A realizes that this packet is a duplicate SYN, that the PCB identifier is different from the one advertized by B. It will refuse to consider the packet's informations, and immediately send a "reset" to C, clearing the additional connection.

In the rare case where the binary values of xb and xc would be identical, A should simply discard the additional asynchronisation packet. It must not add the corresponding source address to its PCB.

4.3. Multi-homing

Suppose that the host A is equipped with two interfaces, A1 and A2. It may decide to spread the traffic over these two interfaces, sending packets alternatively from sources A1 and A2. B, in turn, will send packets alternatively towards each of these addresses:

```
          (A1, B1, tcp) (a, b, SYN, PCBID=xa) ===========>
                 <============ (B1, A1, etcp) (xa, SYN, ACK, PCBID=xb)
          (A1, B1, etcp) (xb, ACK, DATA) ===========>
                 <============ (B1, A1, etcp) (xa, ACK)
          (A2, B1, etcp) (xb, ACK, DATA) ===========>
```


RED HAT 1010
RED HAT V. CAS

```
                <============ (B1, A2, etcp) (xa, DATA)
        (A1, B1, etcp) (xb, ACK, DATA) ===========>
                <============ (B1, A1, etcp) (xa, ACK)
```

This is possible because B keeps in the PCB identified by "xb" a list of partners addresses. Both addresses will in turn be the "most recently seen" address.

4.4. Mobility

Mobility is treated by exactly the same mechanisms as those of "renumbering" which we expressed in 4.1. There are however two problems, fast movements and double jumps. There are few good solutions to the "fast movement" problems. If a host gets a new address each fraction of a second, there is just no way that a mechanism that exchange information with communicating hosts through the whole internet will keep pace. Proposed solutions involve home base and tunnelling. They amount to saying that the mobile host is dual homed. It has a "mobile address", which changes when the host roams, and a "home address", which remains relatively static. When the host moves very fast, it should only use its home address. When the host moves slowly, it can start using the "mobile address" for more efficiency.

The double jump problem occurs when two mobile hosts change cells simultaneously. In this case, we are in a situation were neither of the previous addresses is usable. A cannot tell B that its address changed from A1 to A2, because the packets bound to B1 don't make it to B, and vice versa.

The only solution here is to request that mobile host maintain a stable

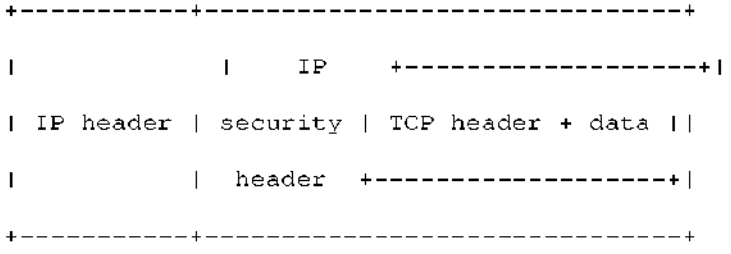
RED HAT 1010
RED HAT V. CAS

home address, e.g. A0 for A and B0 for B. They will "refresh" the address list of their partners by regularly sending messages from these addresses. In the double jump situations, the deadlock will be broken if either A or B manages to get a packet to the home addresses A0 or B0.

5. Security considerations

The procedure that we propose introduces a significant security hole. If a third party can obtain the "context-id" used by the TCP connection, then it can send TCP packets using a bogus source address and trick the TCP entity into believing that this is the new address of its partner: it can, in short, "grab the connection". In the current Internet, such connection grabbings require that one somehow subvert the routing protocol, which is more difficult than simply listening to traffic and forging packets. This threat is in fact inherent to the mobility environment which we are willing to accomodate: we have the choice of accepting this mode of operation or implementing a secure version of TCP.

Implementing a secure version of TCP may be a good idea in any case, as TCP is currently vulnerable to several forms of attacks: as long as the origin of the packet is not securely authenticated, it is easy for intruders to send forged TCP messages and distort the state of an existing connection. We may be willing to use "IP security" to that effect, wrapping the TCP packets inside an "IP secure" payload:



RED HAT 1010
RED HAT V. CAS

The IP security header identifies a "security context", to which a key is attached, as well as the type of protection. Once a security context has been established, we obtain a proof of the source Internet address, as well as in some cases an integrity check, an confidentiality, i.e. protection against eavesdropping. How secure IP contexts are established is outside the scope of these memo.

If such protections are available, then our procedure becomes entirely acceptable. In fact, we can define three mode of operations:

(1) Secure mode, into which only those packets which were transmitted securely can be considered by TCP,

(2) Protected mode, into which packets are requested to come either from a well identified address or through a secure channel,

(3) Unprotected mode, for users who are willing to take chances.

The second mode of operation protects us against the specific threats brought in by mobility, without imposing the performance penalties of cryptography on each and every packet.

We should note that it is theoretically feasible to associate security parameters to the TCP connection, in much the same way that they are associated to a pair of IP hosts. This would solve the TCP problem entirely, making our connection very secure. But this is clearly a point of debate.

8. Acknowledgements and related works

RED HAT 1010
RED HAT V. CAS

Many of the features of this proposal were discussed privately with Lixia Zhang, John Wroclavski, Van Jacobson and Greg Minshall. I received comments from Allison Mankin on a previous version of this proposal. The support of multiple addresses bears some similarities with the "transport context names" proposal of Dave Crocker.

Various authors, notably Steve Deering, have proposed an alternative "context identification" mechanism based on "volatile port numbers": a specific port is assigned to each connexion within the SYN exchange, which is essentially equivalent to our 'context identifier". The scheme that we propose here has the advantage of using larger identifiers: 16 bits does not necessarily allow enough contexts. It also has the advantage of being entirely backward compatible (an additional TCP parameter can be ignored) and of being explicit: alternative addresses can only be used if the peer provided a context identifier.

Appendix A. Relation with other proposals.

Letting TCP connections survive a change of addresses is not a new problem.

Partial solutions have been proposed, which essentially insert a "permanent identifier" layer between the transport control and the routing process. Instead of identifying the transport contexts by a pair of source and destination addresses, one would use this "more permanent" identifiers, often called "end point identifiers" or EID. The addresses may change but the EID will remain constant. In the "classic" version of the EID proposal, these identifiers will be present in each and every

RED HAT 1010
RED HAT V. CAS

packet, in addition to the addresses. Several implementations have been proposed, e.g. as innovative IPng formats. Including EIDs in the IP or IPng header itself has been ruled out during the debates, but one could design an intermediate layer between IP and TCP that would carry those headers. The proposal, however, suffers from three drawbacks:

(1) As EID are likely to have the same size than the IPng addresses, we are considering a sizeable overhead.

(2) EID don't exist yet. Creating yet another address space would implies significant administration efforts.

(3) Disconnecting identification and routing removes whatever security there is in the current IP protocol.

The early SIP and SIPP proposals included a scheme that would alleviate the first concern and eliminate the second one. Instead of creating a new numbering space for EIDs, one would simply specialize one address, e.g. the one that was used for sending the initial synchronization packet of TCP. Most connections will simply go on using that address for their whole duration. If the host moves, or if a new provider is selected, the old address will be still be used as nominal source or destination address, but the packet would be explicitly source routed through the new address. This solution however suffers from the same security problems as the generic EID proposal, and does not entirely cure the "change of provider" problem. If the local network is renumbered, the old addresses ceases to be valid.

In fact, that argument applies to all approaches based on "permanent identifiers": we are exchanging a dependency on the IP address for a

RED HAT 1010
RED HAT V. CAS

dependency on these new identifiers. Proponents of the EID approaches are ready to swear on their most sacred books, or on the head of their favorite dog, that these new numbers will remain valid, unique and stable for eternity. We may note however that there is not a single human construct that ever met this requirement. The case is even more obvious if we focus on network history, where exceptions to the unicity and stability rules have been found for all proposed "unique numbers" that were considered, be they IP addresses, which get renumbered, domain names, which change from time to time, or IEEE-802 "MAC addresses", which can easily be misconfigured.

Rather than yet another gigantic administrative effort, our solution involves only locally asserted numbers, the PCB identifiers.

3. Why TCP

Our solution focuses on TCP. One may well object that TCP is not the only protocol used over the Internet, that the fate- sharing also occurs with other transport protocols, notably over UDP. However, a quick analysis shows that "user
controlled" applications, that use the user datagram protocol, are much less sensitive than those application which relie on the services of a transport layer. This is notably the case of "client server" and "real time" application; this is also the case of "secure" applications.

The most popular building block for "client server" applications is the "stateless RPC". In this model, a client simply sends a request to the server, generally a single UDP datagram. Upon reception of this request, the server executes the required actions and sends back a reply, in most cases a single UDP datagram. No "context" is kept by the server, and the time-span of the transaction is generally quite short, which minimizes

RED HAT 1010
RED HAT V. CAS

the risk of observing a move in the middle of a transaction. At worst, if such a transistion occurs, the client may repeat its request. Indeed, if the server moves, the client will have to find its new location, its new address. But that has to be solved by updating the routing process or the name service; it is not a "transport control" problem. It is true that client-server applications may have to maintain long term associations. But they generally do so by identifying "application level contexts", which are identified by the users and servers names: authentication and access controls ought to be based on user names, not on hardware addresses.

Real time applications also use UDP, generally in conjunction with "multicast transmission". The practical experience have shown however that the multicast transmission is very often achieved through a cascade of relais, and that the source address received in the packets is not always that of the actual source. The "real time transport protocol" includes an identification mechanism by which stations are assigned numbers within multicast groups. Real time applications go to great length to reduce their packet size, using computation intensive compression techniques; an additional header in each and every packet would be entirely unwelcomed.

In fact, one may draw the general conclusion that applications which are using datagram services have to incorporate their own control functions. Identification of the partner, as well as general security measures, are an integral part of these applications. Providing them with a "one size fits all" intermediate layer would probably not help.

Then, one may also observe that even if one were to implement a "transparent" solution, e.g. using some form of stable identifiers, the

RED HAT 1010
RED HAT V. CAS

TCP code would still need to be updated. The error and throughput characteristics of mobile networks are different from those of fixed networks; controlling parallel transmission over multiple paths is not quite the same as controlling a single path. Key points of the implementation, such as the retransmission strategy or the congestion control algorithm would have to be revised in any case.

Appendix B. Managing a set of addresses

We have mentioned the need to maintain a "working set" of partner addresses, and to "grade" each address in that working set. However, the procedure described in the previous section only allows for addition of addresses to the connection, not for their removal. This is because we expect that the working set of addresses will simply be a cache of the "most recently used" address of our partners. TCP entities may use whatever strategy you want to maintain the set of addresses within reasonable limits, e.g. LRU replacement. We will describe here some possible manners of grading the addresses.

The first seen and last seen monitoring will produce a first selection of selectable addresses within the working set. However, all the selected addresses should not necessarily be used. Even if several of them should be used, they need not be all used at the same rate.

Suppose that we have sent different packets to different addresses. If one of these addresses has becomed invalid, packet sent to it will be lost. Indeed, packets sent to valid addresses may also be lost, e.g. if a cosmic ray hits a photodiode at the wrong moment. It is however quite

RED HAT 1010
RED HAT V. CAS

reasonable to relate the selection of addresses with the error control: if a packet sent to one particular address failed, further packets should preferably be sent another acceptable address.

Current TCP implementations try to characterize the network they are using by computing round-trip time estimates and congestion windows. When several addresses are in use, one may well believe that each address will be reached by a different path, and that each of these paths will have different characteristics.

In fact, due to the sequential nature of the acknowledgment process, it would be very difficult to maintain more than one round-trip estimate for the connection. However, it is entirely conceivable to maintain a separate congestion avoidance window for each address in the working set, and to mainatin these windows by running parallel versions of the "slow start" algorithm. The window will be initialised when a new address is added to the working set; it will be increased when a packet sent to that address is acknowledged, it will shrink when a packet sent to that address is lost.

Maintaining a congestion window per active address provides for a rationale way of spreading the traffic over the various addresses. It is clearly not the only way. In fact, this may well open the path for very fruitful researches.

RED HAT 1010
RED HAT V. CAS

An End-to-End Approach to Host Mobility

Alex C. Snoeren and Hari Balakrishnan
MIT Laboratory for Computer Science
Cambridge, MA 02139
{snoeren, hari}@lcs.mit.edu

Abstract

We present the design and implementation of an end-to-end architecture for Internet host mobility using dynamic updates to the Domain Name System (DNS) to track host location. Existing TCP connections are retained using secure and efficient connection migration, enabling established connections to seamlessly negotiate a change in endpoint IP addresses without the need for a third party. Our architecture is secure—name updates are effected via the secure DNS update protocol, while TCP connection migration uses a novel set of *Migrate* options—and provides a pure end-system alternative to routing-based approaches such as Mobile IP.

Mobile IP was designed under the principle that fixed Internet hosts and applications were to remain unmodified and only the underlying IP substrate should change. Our architecture requires no changes to the unicast IP substrate, instead modifying transport protocols and applications at the end hosts. We argue that this is not a hindrance to deployment: rather, in a significant number of cases, it allows for an easier deployment path than Mobile IP, while simultaneously giving better performance. We compare and contrast the strengths of end-to-end and network-layer mobility schemes, and argue that end-to-end schemes are better suited to many common mobile applications. Our performance experiments show that handoff times are governed by TCP migrate latencies, and are on the order of a round-trip time of the communicating peers.

1 Introduction

The proliferation of mobile computing devices and wireless networking products over the past decade has made host and service mobility on the Internet an important problem. Delivering data to a mobile host across a network address change without disrupting existing connections can be tackled by introducing a level of indirection in the routing system. This is the approach taken by Mobile IP [27, 29], which deploys a home agent that intercepts packets destined for a host currently away from its home network, and delivers it to the mobile host via a foreign agent in the foreign network. This approach does not require any changes to the fixed (correspondent) hosts in the Internet, but does require changing the underlying IP substrate to effect this *triangle* routing, and an authentication protocol to ensure that connections are not hijacked by a malicious party. Mobile IP is a "pure" routing solution, a network-layer scheme that requires no changes to any higher layer of the Internet protocol stack.

There are many classes of mobile applications [16]: those where other hosts originate connections to a mobile host and can benefit from both host location and handoff support (e.g., a mobile Web server, mobile telephony); those where the mobile host originates all connections, which do not require host location services but can benefit from handoff support (e.g., mail readers, Web browsers); and those where an application-level retry suffices if the network address changes unexpectedly during a short transaction, which need neither to work well (e.g., DNS resolution). We believe that a good end-to-end architecture for host mobility will support all these modes, and empower applications to make the choice best suited to their needs. Our architecture is motivated by, and meets, this goal. It is an end-to-end approach; no changes to the IP substrate are required.

In our mobility architecture, the decision of whether to support transparent connectivity across network address changes (especially useful for mobile servers) or not (not needed for short client-server transactions) is left to the application. While Mobile IP-style, fully-transparent mobility support is general and sufficient for mobile applications, this generality comes at significant cost, complexity, and performance degradation.

To locate mobile hosts as they change their network attachment point, we take advantage of the widely-deployed Domain Name System (DNS) [20] and its ability to support secure dynamic updates [8, 35]. Because most Internet applications resolve hostnames to an IP address at the beginning of a transaction or connection, this approach is viable for initiating new sessions with mobile hosts. When a host changes its network attachment point (IP address), it sends a secure DNS update to one of the name servers in its home domain updating its current location. The name-to-address mappings for these hosts are uncacheable by other domains, so stale bindings are eliminated.

The ability to support continuous communication during periods of mobility without modifying the IP substrate is a more challenging problem. Because TCP is a connection-oriented reliable protocol,

This research was supported in part by DARPA (Grant No. MDA972-99-1-0014), NTT Corporation, and IBM. Alex C. Snoeren is supported by a National Defense Science and Engineering Graduate (NDSEG) Fellowship.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage, and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MobiCom 2000 08/2000 Boston, MA

© 2000 ACM

RED HAT 1010
RED HAT V. CAS

many TCP applications reasonably expect this service model in the face of losses and transient link failures, route changes, or mobility. The two communicating peers must securely negotiate a change in the underlying network-layer IP address and then seamlessly continue communication. Furthermore, an approach that *requires* either communicating peer to learn about the new network-layer address before a move occurs is untenable because network-layer moves may be quite sudden and unpredictable.

We design a new end-to-end TCP option to support the secure migration of an established TCP connection across an IP address change. Using this option, a TCP peer can suspend an open connection and reactivate it from another IP address, transparent to an application that expects uninterrupted reliable communication with the peer. In this protocol, security is achieved through the use of a connection identifier, or *token*, which may be secured by a shared secret key negotiated through an Elliptic Curve Diffie-Hellman (ECDH) key exchange [36] during initial connection establishment. It requires no third party to authenticate migration requests, thereby allowing the end points to use whatever authentication mechanism they choose to establish a trust relationship. Although we only describe details for TCP migration, we find that this idea is general and can be implemented in a like manner for specific UDP-based protocols such as the Real-time Transport Protocol (RTP) to achieve seamless mobility for those protocols as well.

One way of thinking of our work is in the context of the end-to-end argument [32], which observes that functionality is often best implemented in a higher layer at an end system, where it can be done according to the application's specific requirements. We show that it is possible to implement mobility as an end-to-end service without network-layer support, while providing multiple mobility modes. In this sense, this is akin to applications deciding between UDP and TCP as a transport protocol: many opt for UDP's simplicity and timeliness over TCP's reliability. In the same fashion, applications should be able to select the mobility mode of their choice.

The other significant advantage of handling mobility on an end-to-end basis is that it enables higher layers like TCP and HTTP to learn about mobility and adapt to it. For example, it is a good idea after a network route change to restart TCP transmissions from slow start or a window-halving [13] since the bottleneck might have changed, or adapt the transmitted content to reflect new network conditions. These optimizations can be made naturally if mobility is handled end-to-end, since no extra signalling is needed. Indeed, the large body of work in mobile-aware applications [15, 22, 25] can benefit from our architecture.

Experience with previous end-to-end enhancements such as various TCP options (e.g., SACK [19]), path MTU discovery, HTTP/1.1, etc., has shown that such techniques often meet with less resistance to widespread deployment than changes to the IP substrate. This supports our belief that, in addition to the flexibility it offers, an end-to-end approach may be successfully deployed.

We have implemented this mobility architecture in Linux 2.2 and have conducted several experiments with it. We are encouraged by the ease with which seamless mobility can be achieved, the flexibility it provides, and the lack of performance degradation. Since our scheme does not impose any triangle routing anomalies, end-to-end latency for active connections is better than standard Mobile IP, and similar to Mobile IP with route optimization.

The rest of this paper describes the technical details of our approach. In Section 2, we survey related work in the area of mobility support. We describe our architecture in Section 3, and detail our new Migrate TCP option in Section 4. We discuss the security ramifications of our approach in Section 5 and our implementation and performance results in Section 6. We address some deployment issues in Section 7 and conclude in Section 8.

2 Related work

The problem of Internet host mobility has been approached from many angles in the literature, but they can be classified into two categories. Some techniques attempt to handle host relocation in a completely transparent fashion, hiding any changes in network structure from the end hosts. We term these techniques *network-layer* mobility. By contrast, many other approaches attempt to handle relocation at a higher level in the end host.

2.1 Network-layer mobility

Mobile IP (RFC 2002) [29] is the current IETF standard for supporting mobility on the Internet. It provides transparent support for host mobility by inserting a level of indirection into the routing architecture. By elevating the mobile host's *home address* from its function as an interface identifier to an *end-point identifier* (EID), Mobile IP ensures the delivery of packets destined to a mobile host's home address, independent of the host's physical point of attachment to the Internet, as reflected in its *care-of address*. Mobile IP does this by creating a routing tunnel between a mobile host's home network and its care-of address.

Such routing tunnels need to be implemented with care because advertising explicit host routes into the wide-area routing tables destroys routing scalability. Mobile IP uses a *home agent* physically attached to the mobile host's home network to intercept and tunnel packets to the mobile host. Hence, packets undergo *triangle* routing, which is often longer than the optimal unicast path.

Further compounding the problem is the widespread deployment of ingress filters [9], ratified in February 2000 by the IETF as a "Best Current Practice" to combat denial-of-service attacks. With this mechanism, a router does not forward packets with a source address foreign to the local network, which implies that a packet sent by a mobile host in a foreign network with its source address set to its home address will not be forwarded. The solution to this is to use *reverse tunneling*, which tunnels packets originating at the mobile host first to the host's home agent (using the host's care-of address as a source address), and then from there on to the destination using the home address as the source address. Thus, routing anomalies occur in both directions.

Perkins and Johnson present a route optimization option for Mobile IP to avoid triangle routing [28]. Here, correspondent hosts cache the care-of address of mobile hosts, allowing communication to proceed directly. It requires an authenticated message exchange from the home agent to the correspondent host [26]. The resulting Mobile IP scheme achieves performance almost equivalent to ours, but requires modifications to the end hosts' IP layer[1] as well as the

[1] In fact, the draft allows on-path routers to cache the care-of addresses instead of the end host, but this requires modifying yet another level of infrastructure.

RED HAT 1010
RED HAT V. CAS

infrastructure. In contrast, our approach achieves secure, seamless connection migration without a third-party home agent. It also provides a mobile host the ability to pick a mobility mode based on the needs of its applications.

IPv6 provides native support for multiple simultaneous host addresses, and Mobile IPv6 provides mobility support for IPv6 in much the same fashion as Mobile IP for IPv4. IPv6 extensions allow for the specification of a care-of address, which explicitly separates the role of the EID (the host's canonical IP address) and routing location (the care-of address). Gupta and Reddy propose a similar redirection mechanism for IPv4 through the use of ICMP-like control messages which establish care-of bindings at the end hosts [10].

Mysore and Bharghavan propose an interesting approach to network-layer mobility [23], where each mobile host is issued a permanent Class D IP multicast address that can serve as a unique EID. If multicast were widely deployed, this is a promising approach; because a Class D EID has the benefit of being directly routable by the routing infrastructure, it removes the need for an explicit care-of address. However, this scheme requires a robust, scalable, and efficient multicast infrastructure for a large number of sparse groups.

2.2 Higher-layer methods

The home-agent-based approach has also been applied at the transport layer, as in MSOCKS [18], where connection redirection was achieved using a split-connection proxy.

The general idea of using names as a level-of-indirection to handle object and node mobility is part of computer systems folklore. For some years now, people have talked about using the DNS to effect the level-of-indirection needed to support host mobility, but to our knowledge ours is the first specific and complete architecture that uses the DNS to support Internet host mobility. Recently, Adjie-Winoto *et al.* proposed the integration of name resolution and message routing in an Intentional Naming System to implement a "late binding" option that tracks highly mobile services and nodes [1], and it seems possible to improve the performance of that scheme using our connection migration approach.

Our approach differs fundamentally from EID/locator techniques since it requires no additional level of global addressing or indirection, but only a (normally pre-existing) DNS entry and a shared connection key between the two end hosts. Furthermore, unlike previous connection-ID draft proposals such as Huitema's ETCP [11] for TCP connection re-addressing, it requires no modification to the TCP header, packet format, or semantics.[2] Instead, it uses an additional TCP option and the inserts an additional field into the Transmission Control Block (TCB).

There is a large body of work relating to improving TCP performance in wireless and mobile environments [5, 6]. While not the focus of our work, our adherence to standard TCP semantics allows these schemes to continue to work well in our architecture. Furthermore, since end hosts are explicitly notified of mobility, significant performance enhancements can be achieved at the application level [25].

[2]Special RST handling is required on some networks that may rapidly reassign IP addresses; Section 4.5 discusses this issue.

3 An end-to-end architecture

In this section, we describe our end-system mobility architecture. There are three important components in this system: addressing, mobile host location, and connection migration. By giving the mobile host explicit control over its mobility mode, we remove the need for an additional (third-party) home-agent to broker packet routing. The DNS already provides a host location service, and any further control is managed by the communicating peers themselves, triggered by the mobile host when it changes network location.

We assume, like most mobility schemes, that mobile hosts do not change IP addresses more than a few times a minute. We believe this is a reasonable assumption for most common cases of mobility. We emphasize that this does not preclude physical mobility at rapid velocities across a homogeneous link technology, since that can be handled at the physical and link layers, e.g., via link-layer bridging [12].

The rest of this section discusses addressing in a foreign network and host location using the DNS. Section 4 is devoted to a detailed description of TCP connection migration.

3.1 Addressing

The key to the scalability of the Internet architecture is that the IP address serves as a routing locator, reflecting the addressee's point of attachment in the network topology. This enables aggregation based on address prefixes and allows routing to scale well. Our mobility architecture explicitly preserves this crucial property of Internet addressing.

Like Mobile IP, we separate the issues of obtaining an IP address in a foreign domain from locating and seamlessly communicating with mobile hosts. Any suitable mechanism for address allocation may be employed, such as manual assignment, the Dynamic Host Configuration Protocol (DHCP) [7], or an autoconfiguration protocol [34].

While IP addresses fundamentally denote a point of attachment in the Internet topology and say nothing about the identity of the host that may be connected to that attachment point, they have implicitly become associated with other properties as well. For example, they are often used to specify security and access policies as in the case of ingress filtering to alleviate denial-of-service attacks. Our architecture works without violating this trust model and does not require any form of forward or reverse tunneling to maintain seamless connectivity. In a foreign network, a mobile host uses a locally obtained interface address valid in the foreign domain as its source address while communicating with other Internet hosts.

3.2 Locating a mobile host

Once a mobile host obtains an IP address, there are two ways in which it can communicate with correspondent hosts. First, as a client, when it actively opens connections to the correspondent host. In this case, there is no special host location task to be performed in our architecture: using the DNS as before works. However, if the mobile host were to move to another network attachment point during a connection, a new address would be obtained as described in the previous section, and the current connection would continue seamlessly via a secure negotiation with the communicating peer as

RED HAT 1010
RED HAT V. CAS

described in Section 4. If a mobile host were always a client (not an uncommon case today), then no updates need to be made to any third party such as a home agent or the DNS.

To support mobile servers and other applications where Internet hosts actively originate communication with a mobile host, we use the DNS to provide a level of indirection between a host's current location and an invariant end-point identifier. In Mobile IP, a host's home address is the invariant, and all routing (in the absence of route optimization) occurs via the home agent that intercepts packets destined to this invariant. Ours is not a network-layer solution and we can therefore avoid the indirection for every packet transmission. We take advantage of the fact that a hostname lookup is ubiquitously done by most applications that originate communication with a network host, and use the DNS name as the invariant. We believe that this is a good architectural model: a DNS name identifies a host and does not assume anything about the network attachment point to which it may currently be attached, and the indirection occurs only when the initial lookup is done via a control message (a DNS lookup).

This implies that when the mobile host changes its attachment point, it must detect this and change the hostname-to-address ("A-record") mapping in the DNS. Fortunately, both tasks are easy to accomplish, the former by using a user-level daemon as in Mobile IP, and the latter by using the well-understood and widely available secure DNS update protocol [8, 35]. We note that some DHCP servers today issue a DNS update at client boot time when handing out a new address to a known client based on a static MAC-to-DNS table. This augurs well for the incremental deployability of our architecture, since DNS update support is widely available.

The DNS provides a mechanism by which name resolvers can cache name mappings for some period of time, specified in the time-to-live (TTL) field of the A-record. To avoid a stale mapping from being used from the name cache, we set the time-to-live (TTL) field for the A-record of the name of the mobile host to *zero*, which prevents this from being cached.[3] Contrary to what some might expect, this does not cause a significant scaling problem; name lookups for an uncached A-record do not have to start from a root name server, because in general the "NS-record" (name server record) of the mobile host's DNS name is cacheable for a long period of time (many hours by default). This causes the name lookup to start at the name server of the mobile host's domain, which scales well because of administrative delegation of the namespace and DNS server replication in any domain. We note that some content distribution networks for Web server replication of popular sites use the same approach of small-to-zero TTL values to redirect client requests to appropriate servers (e.g., Akamai [2]). There is no central hot spot because the name server records for a domain are themselves cacheable for relatively long periods of time.

Even with uncacheable DNS entries there still exists a possible race condition where a mobile host moves between when a correspondent host receives the result of its DNS query and when it initiates a TCP connection. Assuming a mobile host updates its DNS entry immediately upon reconnection, the chances of such an occurrence are quite small, but non-zero, especially for a mobile host that makes frequent moves. In this case, the correspondent host will attempt to open a TCP connection to the mobile host's old address, and has no automatic fail-over mechanism.

In this case, the application must perform another DNS lookup to find the new location of the mobile host. We note that the trend towards dynamic DNS records has caused such application-level retries to find their way into applications already—for instance, current FreeBSD `telnet` and `rsh` applications try alternate addresses if an initial connection fails to a host that has multiple DNS A-records. It seems to be only a minor addition to refresh DNS bindings if connection establishment fails.

4 TCP connection migration

A TCP connection [31] is uniquely identified by a 4-tuple: ⟨*source address*, *source port*, *dest address*, *dest port*⟩. Packets addressed to a different address, even if successfully delivered to the TCP stack on the mobile host, must not be demultiplexed to a connection established from a different address. Similarly, packets from a new address are also not associated with connections established from a previous address. This is crucial to the proper operation of servers on well-known ports.

We propose a new *Migrate* TCP option, included in SYN segments, that identifies a SYN packet as part of a previously established connection, rather than a request for a new connection. This Migrate option contains a *token* that identifies a previously established connection on the same destination ⟨*address*, *port*⟩ pair. The token is negotiated during initial connection establishment through the use of a *Migrate-Permitted* option. After a successful token negotiation, TCP connections may be uniquely identified by either their traditional ⟨*source address*, *source port*, *dest address*, *dest port*⟩ 4-tuple, or a new ⟨*source address*, *source port*, *token*⟩ triple on each host.

A mobile host may restart a previously-established TCP connection from a new address by sending a special Migrate SYN packet that contains the token identifying the previous connection. The fixed host will than re-synchronize the connection with the mobile host at the new end point. A migrated connection maintains the same control block and state (with a different end point, of course), including the sequence number space, so any necessary retransmissions can be requested in the standard fashion. This also ensures that SACK and any similar options continue to operate properly. Furthermore, any options negotiated on the initial SYN exchange remain in effect after connection migration, and need not be resent in a Migrate SYN.[4]

Since SYN segments consume a byte in the TCP sequence number space, Migrate SYNs are issued with the same sequence number as the last transmitted byte of data. This results in two bytes of data in a migrated TCP connection with the same sequence number (the new SYN and the previously-transmitted actual data), but this is not a problem since the Migrate SYN segment need never be explicitly acknowledged. Any packet received from the fixed host by a migrating host at the mobile host's new address that has a sequence number in the appropriate window for the current connection implicitly acknowledges the Migrate SYN. Similarly, any further seg-

[3] Modern versions of BIND honor this correctly.

[4] They can be, if needed. For example, it might be useful to renegotiate a new maximum segment size (MSS) reflecting the properties of the new path. We have not yet explored this in detail.

RED HAT 1010
RED HAT V. CAS

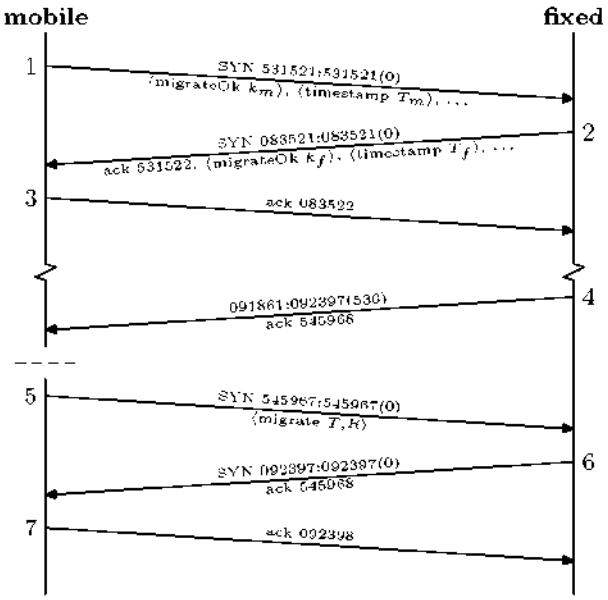


Figure 1: TCP Connection Migration

ments from the mobile host provide the fixed host an implicit acknowledgement of its SYN/ACK. Thus, there is exactly one byte in the sequence space that needs explicit acknowledgement even when the Migrate SYN is used.

4.1 An example

Figure 1 shows a sample connection where a mobile client connects to a fixed host and later moves to a new address. The mobile client initiates the TCP connection in standard fashion in message 1, including a Migrate-Permitted option in the SYN packet. The values k_m and T_m are parameters used in the token negotiation, described in Section 4.3. The fixed server, with a migrate-compliant TCP stack, indicates its acceptance of the Migrate-Permitted option by including the Migrate-Permitted option in its response (message 2). The client completes the three-way handshake with message 3, an ACK. The connection then proceeds as any other TCP connection would, until message 4, the last packet from the fixed host to the mobile host at its current address.

At some time later the mobile host moves to a new address, and notifies the fixed server by sending a SYN packet from its new address in message 5. This SYN includes the Migrate option, which contains the previously computed connection token as part of a migration request. Note that the sequence number of this Migrate SYN segment is the same as the last byte of transmitted data. The server responds in kind in message 6, also using the sequence number of its last transmitted byte of data. The ACK, however, is from the same sequence space as the previous connection. While in this example it acknowledges the same sequence number as the SYN that generated it, it could be the case that segments were lost during a period of disconnect while the mobile host moves, and that the ACK will be a duplicate ACK for the last successfully received in-sequence byte. Since it is addressed to the mobile host's new location, however, it serves as an implicit ACK of the SYN as well.

Upon receipt of this SYN/ACK, the mobile host similarly ACKs in the previous sequence space, and the connection resumes as before. All of the options negotiated on the initial SYN except the Migrate-Permitted option are still in effect, and need not be replicated in this or any subsequent migrations.

4.2 Securing the migration

It is possible to partially hijack TCP connections if an attacker can guess the sequence space being used by the connection [21]. With the Migrate options, an attacker who can guess both the sequence space and the connection token can hijack the connection completely. Furthermore, the ability to generate a Migrate SYN from anywhere greatly increases the connection's exposure. While ingress filtering can be used to prevent connection hijacking by attackers not on the path between the end hosts, such methods are ineffective in our case. We must therefore take care to secure the connection token.

The problem is relatively easy to solve if IP security (IPsec) [4] were deployed. While the spectrum of approaches that could be used is outside the scope of this paper, we note that IPsec provides sufficient mechanisms to secure migrateable connections. Currently, however, IPsec has not found wide-spread deployment. Hence, we provide a mechanism to self-secure the Migrate options. End hosts may elect to secretly negotiate an unguessable connection token, which then reduces the security of a migrateable TCP connection to that of a standard TCP connection, since no additional attacks are possible against a migrateable connection without guessing the token, and any attack against a standard TCP connection clearly remains feasible against a migrateable TCP connection.

An unguessable connection token is secured with a secret *connection key*. Since any host that obtains the connection key could fabricate the token and issue a Migrate request, we select the key with an Elliptic Curve Diffie-Hellman key exchange [36], as described below. Hosts using IPsec, or unconcerned with connection security, may choose to disable key negotiation to avoid excess computation.

4.3 Migrate-Permitted option

Hosts wishing to initiate a migrateable TCP connection send a Migrate-Permitted option in the initial SYN segment. Similar to the SACK-Permitted option [19], it should only be sent on SYN segments, and not during an established connection. Additionally, hosts wishing to cryptographically secure the connection token may conduct an Elliptic Curve Diffie-Hellman (ECDH) key exchange through the option negotiation. (Elliptic Curve Diffie-Hellman is preferred to other methods of key establishment due to its high security-to-bit-length ratio. Readers unfamiliar with Elliptic Curve cryptography can find the necessary background material in [3].)

As seen in figure 2, the Migrate-Permitted option comes in two variants—the insecure version, of length 3, and the secure version, with length 20. The secure version is used to negotiate a secret connection key, and contains an 8-bit *Curve Name* and a 136-bit *ECDH Public Key* fragment. The curve name field selects a particular set of domain parameters (the curve, underlying finite field, F, and its representation, the generating point, P, and the order of P, n), as specified in [3]. Use of the insecure version, which contains only a Curve Name field (which must be set to zero) allows the end host

RED HAT 1010
RED HAT V. CAS

Kind: 15	Length = 3/20	Curve Name	ECDH PK
ECDH Public Key (cont.)			
ECDH Public Key (cont.)			
ECDH Public Key (cont.)			
ECDH Public Key (cont.)			

Figure 2: TCP Migrate-Permitted option

	Kind: 16	Length = 19	ReqNo
Token			
Token (cont.)			
Request			
Request (cont.)			

Figure 3: TCP Migrate option

to skip the key negotiation process. In that case, the connection key is set to all zeros.

The secure variant of the Migrate-Permitted option also requires the use of the Timestamp [14] option in order to store up to 200 bits of ECDH keying material. The EDCH Public Key is encoded using the compressed conversion routine described in [3, Section 4.3.6]. The 136 least-significant bits are stored in the EDCH Public Key field of the Migrate-Permitted option, while the remaining 64 bits of the key are encoded in the Timestamp option. The timestamp option, while often included, is not used on SYN segments. The Protection Against Wrapped Sequence Numbers (PAWS) [14] check is only performed on synchronized connections, which by definition [31] includes only segments after the three-way handshake. Similarly, the Round-Trip Time Measurement (RTTM) [14] procedure only functions when a timestamp has been echoed—clearly this is never the case on an initial SYN segment. Hence the value of the Timestamp option on SYN segments is entirely irrelevant to current TCP stacks. Legacy TCP stacks will never receive a Migrate-Permitted option on a SYN/ACK, hence the Timestamp option will be processed normally. Special handling is only required for the SYN/ACK and following ACK segment on connections that have negotiated the Migrate-Permitted option, as Timestamp fields on these segments will not contain timestamps. Hence the RTTM algorithm must not be invoked for SYN/ACK or initial ACK segments of connections that have negotiated the Migrate-Permitted option.

The Timestamp *TSVal* field contains the 32 most-significant bits of the public key, while the *TSecr* field contains the next 32 most-significant bits. These two components, combined with the 136-bit EDCH Public Key field of the Migrate-Permitted option, constitute the host's public key, k. If the public key is less than 200 bits, it is left-padded with zeros. For any host, i, k_i is generated by selecting a random number, $X_i \in [1, n-1]$, where n is the order of P, and computing

$$k_i = X_i * P$$

The $*$ operation is the scalar multiplication operation over the field F. The security of the connection hinges on the secrecy of the negotiated key, hence X_i should be randomly generated and stored in the control block for each new connection. Any necessary retransmissions of the SYN or SYN/ACK must include identical values for the Migrate-Permitted and Timestamp option.

Upon receipt of an initial SYN with a Migrate-Permitted option, a host, j, with a compliant TCP stack must include a Migrate-Permitted option (and a Timestamp option if the secure variant is used) in its SYN/ACK segment. It similarly selects a random $X_j \in [1, n-1]$ which it uses to construct k_j, its public key, which it sends in the same fashion.

After the initiating host's reception of the SYN/ACK with the Migrate-Permitted and Timestamp options, both hosts can then compute a shared secret key, K, as specified in [36]:

$$K = k_i * X_j = k_j * X_i$$

This secret key is then used to compute a connection validation token. This token, T, is computed by hashing together the key and the initial sequence numbers N_i and N_j using the Secure Hash Algorithm (SHA-1) [24] in the following fashion (recall that host i initiated the connection with an active open, and host j is performing a passive open):

$$T = SHA1(N_i, N_j, K)$$

While SHA-1 produces a 160-bit hash, all but the 64 most-significant bits are discarded, resulting in a cryptographically-secure 64-bit token that is unique to the particular connection. Since SHA-1 is collision-resistant, the chance that another connection on the same $\langle address, port \rangle$ pair has an identical token is extremely unlikely. If a collision is detected, however, the connection must be aborted by sending a RST segment. (The host performing a passive open can check for collisions before issuing a SYN/ACK, and select a new random X_j until a unique token is obtained. Hence the only chance of collision occurs on the host performing the active open.)

4.4 Migrate option

The Migrate option is used to request the migration of a currently open TCP connection to a new address. It is sent in a SYN segment to a host with which a previously-established connection already exists (in the ESTABLISHED or FIN_WAIT states), over which the Migrate-Permitted option has been negotiated.

There are two 64-bit fields in a Migrate option: a *token*, and a *request*. In addition, there is an 8-bit sequence number field, *reqNo*, which must be monotonically increasing with each new migrate request issued by an end host for a connection. (The sequence number allows correspondent hosts to ensure Migrate SYNs were not reordered by the network. Sequence space wrap-around is dealt with in the standard fashion.) The token is simply the 64 most-significant bits of the connection's SHA-1 hash as computed in the Migrate-Permitted option exchange. The request, R, is similarly

RED HAT 1010
RED HAT V. CAS

the 64 most-significant bits of a SHA-1 hash calculated from the sequence number of the connection initial sequence numbers N, Migrate SYN segment, S, the connection key, K, and the request sequence number, I.

$$R - SHA1(N_i, N_j, K, S, I)$$

SYN segments may now correctly arrive on a bound port not in the LISTEN state. They should be processed only if they contain the Migrate option as specified above. Otherwise, they should be treated as specified in [31]. Upon receipt of a SYN packet with the Migrate option, a TCP stack that supports migration attempts to locate the connection on the receiving port with the corresponding token. The token values for each connection were precomputed at connection establishment, reducing the search to a hash lookup.

If the token is valid, meaning an established connection on this ⟨*address*, *port*⟩ pair has the same token, and the *reqNo* is greater than any previously received migrate request, the fixed host then computes $R = SHA1(N_i, N_j, K, S, I)$ as described above, and compares it with the value of the request in the Migrate SYN. If the comparison fails, or the token was invalid, a RST is sent to the address and port issuing the Migrate SYN, and the SYN ignored. If, on the other hand, the token and request are valid, but the reqNo is smaller than a previously received request, the SYN is assumed to be out-of-order and silently discarded. If the reqNo is identical to the most recently received migrate request this SYN is assumed to be a duplicate of the most recently received SYN, and processed accordingly.

Otherwise, the destination address and port[5] associated with the matching connection should be updated to reflect the source of the Migrate SYN, and a SYN/ACK packet generated, with the ACK field set to the last received contiguous byte of data, and the connection placed in the SYN_RCVD state. Upon receipt of an ACK, the connection continues as before.

4.5 MIGRATE_WAIT state

This section assumes that the reader is familiar with the TCP state machine and transitions [33, Chapter 18].

Special processing of TCP RST messages is required with migrateable connections, as a mobile host's old IP address may be reassigned before it has issued a migrate request to the fixed host. Figure 4 shows the modified TCP state transition diagram for connections that have successfully negotiated the Migrate-Permitted option. The receipt of a RST that passes the standard sequence number checks in the ESTABLISHED state does not immediately terminate the connection, as specified in [31]. Instead, the connection is placed into a new *MIGRATE_WAIT* state. (A similar, but far less likely situation can occur if the fixed host is in the FIN_WAIT1 state—the application on the fixed host has closed the connection, but there remains data in the connection buffer to be transmitted. For simplicity, these additional state transitions are not shown in figure 4.)

Connections in the MIGRATE_WAIT state function as if they were in the ESTABLISHED state, except that they do not emit any segments (data or ACKs), and are moved to CLOSED if they remain

[5]Migrated connections will generally originate from the same port as before. However, if the mobile host is behind a NAT, it is possible the connection has been mapped to a different port.

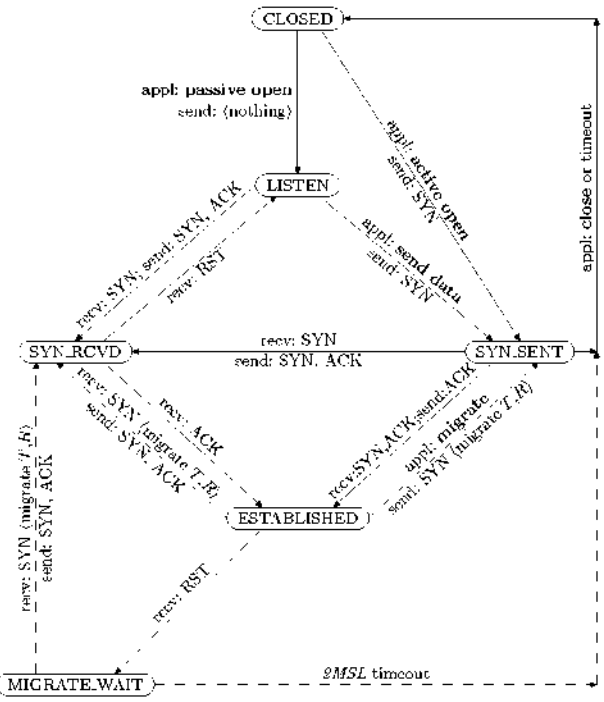


Figure 4: Partial TCP state transition diagram with Migrate transitions (adapted from [33, figure 18.12])

in MIGRATE_WAIT for over a specified period of time. We recommend using the *2MSL* ([31] specifies a Maximum Segment Lifetime (MSL) as 2 minutes, but common implementations also use values of 1 minute or 30 seconds for MSL [33]) period of time specified for the TIME_WAIT state.

Any segments received while in the MIGRATE_WAIT state should be processed as in the ESTABLISHED state, except that no ACKs should be generated. The only way a connection is removed from the MIGRATE_WAIT state is on the receipt of a Migrate SYN with the corresponding connection key. The connection then responds in the same fashion as if it were in the ESTABLISHED state when it received the SYN.

The MIGRATE_WAIT state prevents connections from being inadvertently dropped if the address allocation policy on the mobile host's previous network reassigns the mobile host's old IP address before the mobile host has reconnected at a new location and had a chance to migrate the connection. It also prevents the continued retransmission of data to an unreachable host.

This passive approach to disconnection discovery is preferred over an active, mobile-initiated squelch message because any such message could be lost.[6] Furthermore, a mobile host may not have sufficient (if any) notice of address reassignment to issue such messages. As an added performance enhancement, however, mobile hosts aware of an impending migration may themselves emit a special RST to the peer, which will force the connection into MIGRATE_WAIT, preventing additional packet transmission until the

[6]And any guaranteed-reliable transmission mechanism could take unbounded time.

RED HAT 1010
RED HAT V. CAS

mobile host has successfully relocated, although such action invokes the strict 2MSL time bound on the allowable delay for host relocation and connection migration.

5 Security issues

An end-to-end approach to mobility simplifies the trust relationships required to securely support end-host mobility compared to network-layer approaches such as Mobile IP. In addition to the relationship between a mobile host and any proxies or home agents, several Mobile IP-based proposals require that a correspondent host in communication with a mobile host assume the responsibility of authenticating communication with an arbitrary set of foreign agents. In their route optimization draft [28], Perkins and Johnson state:

> One of the most difficult aspects of Route Optimization for Mobile IP in the Internet today is that of providing authentication for all messages that affect the routing of datagrams to a mobile node.

Since no third parties are required or even authorized to speak on the mobile host's behalf in an end-to-end architecture, the only trust relationship required for secure relocation is between the mobile and correspondent host. Clearly they already must have a level of trust commensurate with the nature of their communications since they chose to communicate in the first place.

Regardless of the simplicity of trust relationships, there remains the possibility that untrusted parties could launch attacks against the end hosts or connections between them utilizing either dynamic DNS updates or the Migrate and Migrate-Permitted options. The security of dynamic DNS updates is addressed in RFC 2137 [8], resting on the strength of the digital signature scheme used to authenticate mobile hosts.

Possible attacks against the Migrate TCP options include both denial-of-service attacks and methods of migrating connections away from their appropriate end hosts. We discuss these attacks below, and either show why the Migrate options are not vulnerable, or explain why the attack presents no additional threat in relation to standard TCP.

5.1 Denial of service

SYN flooding is a common form of Denial-of-Service (DoS) attack, and most modern TCP implementations have taken great care to avoid consuming unnecessary resources unless a three-way handshake is complete. To validate a Migrate request, the correspondent host performs a significant computation (the SHA-1 hash), which implies we need to be especially vigilant against DoS attacks that attempt to deplete the CPU resources of a target host. The validation is not performed unless an attacker succeeds in guessing a valid, pre-computable token (with a 1 in 2^{64} probability); since a RST message is generated if either the token or the request is invalid, an attacker has no way to identify when it has found a valid token. Because a would-be attacker would therefore have to issue roughly 2^{63} Migrate SYNs to force a request validation, we argue that the TCP Migrate option does not introduce any additional DoS concerns above standard TCP.

5.2 Connection hijacking

Since a Migrate request contains a hash of both the SYN segment's sequence number and migrate request sequence number, a replayed Migrate option can only be used until either a new byte of data or another migrate connection is sent on the connection. Since self-migration is not allowed, duplicate Migrate SYNs (received outside of the three-way handshake) are ignored by the peer TCP. If, however, the mobile host moves rapidly to a another new location, a replayed Migrate SYN could be used to migrate the connection back to the mobile host's previous IP, which may have been subsequently assumed by the attacker. In order to prevent this attack, the Migrate Request option processing ignores the source address and port in duplicate packets, as a valid request from a relocated mobile host would include a higher request number.

More worrisome, however, is the fact that once a Migrate SYN has been transmitted, the token is known by any hosts on the new path, and denial-of-service attacks could be launched by sending bogus Migrate SYNs with valid tokens. If a mobile host includes a new Migrate-Permitted option in its Migrate SYN, however, the window of opportunity when the previous connection token can be used (if it was snooped) is quite small—only until the new three-way handshake is successfully completed.

5.3 Key security

The connection key used by the Migrate option is negotiated via Elliptic Curve Diffie-Hellman to make it extremely difficult even for hosts that can eavesdrop on the connection in both directions to guess the key. Without sufficient information to verify possible keys off-line, an attacker would have to continually generate Migrate SYNs and transmit them to one of the end hosts, hoping to receive a SYN/ACK in response to a correct guess. Clearly such an attack is of little concern in practice, as the expected 2^{63} SYN packets required to successfully guess the key would generate sufficient load as to be a DoS problem in and of themselves.

Hosts that lie on the path between end hosts, however, have sufficient information (namely the two Elliptic Curve Diffie-Hellman components) to launch an attack against the Elliptic Curve system itself. The best known attack is a distributed version of Pollard's rho-algorithm [30], which [17] uses to show that a 193-bit EC system would require $8.52 \cdot 10^{14}$ MIPS years, or about $1.89 \cdot 10^{12}$ years on a 450Mhz Pentium II, to defeat.

While this seems more than secure against ordinary attackers, an extremely well-financed attacker might be able to launch such an attack on a long-running connection in the not-too-distant future. The obvious response is to increase the key space. Unfortunately, we are restricted by the 40-byte limitation on TCP options. Given the prevalence of the MSS (4 bytes), Window Scale (3 bytes), SACK Permitted (2 bytes), and Timestamp (10 bytes) options (of which we are already using 8 bytes) in today's SYN segments, the 20-byte Migrate-Permitted option is already as large as is feasible. We argue that further securing the connection key against brute-force attacks from hosts on the path between the two end hosts is largely irrelevant, given the ability of such hosts to launch man-in-the-middle attacks against TCP with much less difficulty!

The security of TCP connections, migrateable or not, continues to remain with the authentication of end hosts, and the establishment

RED HAT 1010
RED HAT V. CAS

of strong session keys to authenticate ongoing communication. Although we have taken care to ensure the Migrate option does not further decrease the security of TCP connections, the latter are inherently insecure, as IP address spoofing and sequence number guessing are not very difficult. Hence we strongly caution users concerned with connection security to use additional application-layer cryptographic techniques to authenticate end points and the payload traffic.

5.4 IPsec

When used in conjunction with IPsec [4], there are additional issues raised by the use of the Migrate options. IPsec Security Associations (SAs) are established on an IP-address basis. When a connection with an associated SA is migrated, a new SA must be established with the new destination address before communication is resumed. If the establishment of a this new SA conflicts with existing policy, the connection is dropped. This seemingly unfortunate result is actually appropriate. Since IPsec's Security Policy Database (SPD) is keyed on IP network address, the policies specified within speak to a belief about the trustworthiness of a particular portion of the network.

If a mobile host attaches to a foreign network, any security assumptions based on its normal point of attachment are invalid. If the end host itself continues to have sufficient credentials independent of its point of attachment, an end-to-end authentication method should be used, and a secure tunnel established for communication over the untrusted network. A discussion of such techniques is outside of the scope of this document.

6 Implementation

We have implemented this architecture in the Linux 2.2.15 kernel, using Bind 8.2.2-P3 as the name server for mobile hosts. The IPv4 TCP stack has been modified to support the Migrate options. Connection migration can be affected through two methods. Applications with open connections may explicitly request a migration by issuing an `ioctl()` on the connection's file descriptor specifying the address to migrate to. Most current applications, however, lack a notification method so the system can inform them the host has moved. Hence we also provide a mechanism for processes to migrate open connections, regardless of whether they have the file descriptor open or not.

This is done through the Linux `/proc` file system. A directory `/proc/net/migrate` contains files of the form *source address*`:`*source port*`->`*dest address*`:`*dest port* for each open connection that has successfully negotiated the Migrate-Permitted option. These files are owned by the user associated with the process that opened the connection. Any process with appropriate permissions can then write a new IP address to these files, causing the corresponding connection to be migrated to the specified address. This method has the added benefit of being readily accessed by a user directly through the command line.

It is expected that mobile hosts will run a mobility daemon that tracks current points of network attachment, and migrates open connections based on some policy about the user's preference for certain methods of attachment. For instance, when an 802.11 interface comes up on a laptop that previously established connections on a CDPD link, it seems likely that the user would opt to migrate most open connections to the address associated with the 802.11 link. Similarly the daemon could watch for address changes on attached interfaces (possibly as a result of DHCP lease expirations and renewals) and migrate connections appropriately. We plan to implement such a daemon in the near future.

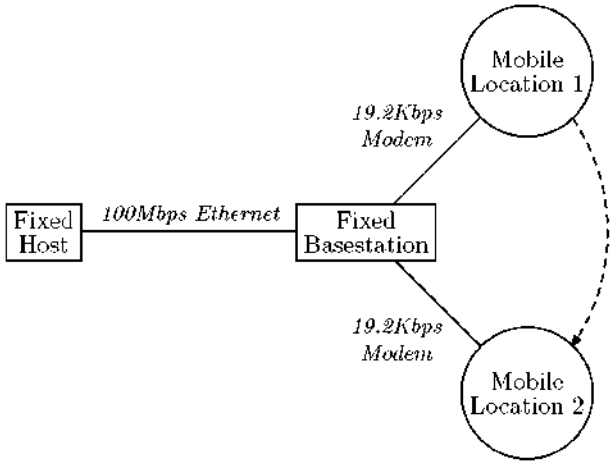


Figure 5: Network topology used for migration experiments

6.1 Experiments

Figure 5 shows the network topology used to gather the TCP traces shown in figures 6 and 7. The traces were collected at the fixed basestation, which is on the path between the fixed host and both mobile host locations. We conducted TCP bulk transfers from a server on the fixed host to a client on the mobile host. The client initiates the connection from one location, and migrates to another location at some later point. Both mobile host locations use identical connections, a 19.2Kbps serial link with ≈100ms round-trip latency. The basestation and fixed host are on a 100Mbps Ethernet segment, hence the link to the mobile host is the connection bottleneck. This topology is intentionally simple in order to isolate the various subtleties of migrating TCP connections, as discussed below.

Figure 6 shows the TCP sequence trace of a migrated TCP connection. At time $t \approx 4.0s$ the mobile host moved to a new address and issued a Migrate SYN, as depicted by the dotted line. Since the host is no longer attached at its previous address, all of the enqueued segments at the bottleneck are lost. (The amount of lost data is bounded by the advertised receive window of the mobile host. A host that moves frequently across low-bandwidth connections may wish to advertise a smaller receive window to reduce the number of wasted segments.) Finally, at $t \approx 6.8s$ the fixed host's SYN/ACK passes through the bottleneck, and is ACKed by the fixed host a RTT later.

The fixed host does not immediately restart data transmissions because the TCP Migrate options do not change the congestion-avoidance or retransmission behavior of TCP. The sender is still waiting for ACKs for the lost segments; as far as it is concerned, it has only received two (identical) ACKs—the original ACK, and one duplicate as part of the Migrate SYN three-way handshake.

RED HAT 1010
RED HAT V. CAS

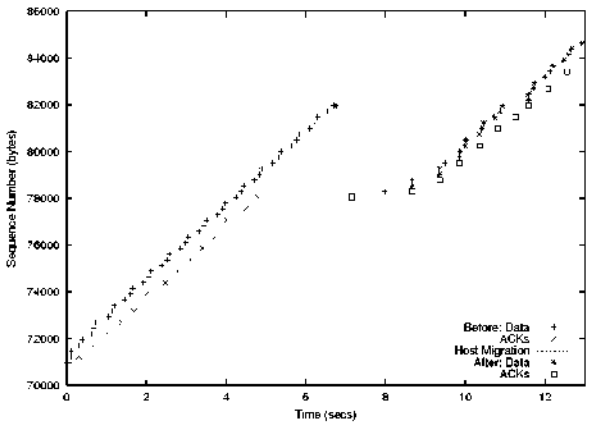


Figure 6: A TCP connection sequence trace showing the migration of an open connection

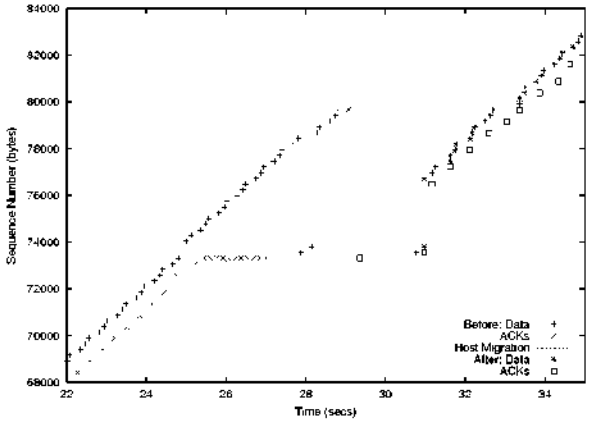


Figure 7: A TCP Migrate connection (with SACK) sequence trace with losses just before migration

Finally, at $t \approx 7.8s$ the retransmission timer expires (the interval is from the first ACK, sent earlier at $t \approx 4.9s$) and the fixed host retransmits the first of the lost segments. It is immediately acknowledged by the mobile host, and TCP resumes transmission in slow-start after the timeout.

Figure 7 shows the TCP sequence trace of a similar migrate TCP connection. As before, the dashed line indicates the mobile host issued a migrate request at time $t \approx 27.1s$. This time, however, there were additional losses on the connection that occurred just before the migration, as can be seen at $t \approx 24.9s$. These segments are fast-retransmitted, and pass through the bottleneck at $t \approx 28s$ due to the DUP-ACKs generated by the remaining SYNs. Unfortunately, this is after the mobile host has migrated, so they, along with all the segments addressed to the mobile host's initial address after $t \approx 27.1s$, are lost.

At $t \approx 29s$, the Migrate SYN/ACK makes it out of the queue at the bottleneck, and the mobile host immediately generates an ACK. As in the previous example, however, the fixed host is still awaiting ACKs for previously transmitted segments. It is only at $t \approx 31s$ that the timer expires and the missing segments are retransmitted. Notice that because SACK prevents the retransmission of the previously-received segments, only those segments lost due to the mobile host's address change are retransmitted, and the connection continues as before. The success of this trace demonstrates that the Migrate options work well with SACK due to the consistency of the sequence space across migrations.

6.2 Performance enhancements

Several enhancements can be made by implementations to improve overall connection throughput during connection migration. The most obvious of these is issuing three DUP-ACKs immediately after a migrate request, thereby triggering the fast-retransmit algorithm and avoiding the timeout seen in the previous example [6]. By preempting the timeout, the connection further avoids dropping into slow-start and congestion avoidance.

Such techniques should be used with care, however, as they assume the available bandwidth of the new path between mobile and fixed host is on the same order-of-magnitude as the previous path. For migrations across homogeneous technologies this may be a reasonable assumption. When moving from local to wide-area technologies, however, there may be order-of-magnitude discrepancies in the available bandwidth. Hence we do not include such speed-ups in the TCP Migrate specification, and leave it to particular implementations to responsibly evaluate the circumstances and provide behavior compatible with standard TCP.

7 Deployment Issues

As with any scheme for mobility support, there are some deployment issues to be addressed. By pushing the implementation of mobility mechanisms—connection migration in particular—to the end points, our system requires changes to each transport protocol. Fortunately, our TCP connection migration protocol can be generalized to other UDP-based protocols with little difficulty. Significant examples include streaming protocols such as RTP and proprietary protocols like Real, Quicktime and Netshow. We note that most of these already have a control channel used for congestion and quality control, and such applications would likely wish to be informed of changes due to mobility as well. Furthermore, we argue that not all applications *require* network-layer mobility, especially those characterized by short transactions where an application-level retry of the transaction is easy to perform; we therefore make the case using the end-to-end argument that mobility might be best implemented as a higher-level, end-to-end function just like reliability.

Perhaps the biggest limitation of our approach is that both peers cannot move *simultaneously*.[7] Because our scheme has no anchor point like Mobile IP's home agent, any IP address change must be completed before the other can proceed. We do not view this as a serious limitation to the widespread applicability of the protocol, since we are primarily targeting infrastructure-based rather than ad-hoc network topologies in this work.

In addition to these two limitations, there are several issues that crop up when one considers presently-deployed applications. While it is currently possible for Internet hosts to be re-addressed while

[7] "Simultaneously" is defined as whenever the intervals between address change and the (would-be) reception of the Migrate SYN by the corresponding host for both end hosts overlap.

RED HAT 1010
RED HAT V. CAS

operating (due to a DHCP lease expiration or similar event), it is quite rare. Hence some applications have made assumptions about the stability of network addresses, which are no longer valid in our architecture. We discuss some of these issues below.

7.1 Address caching

There is a class of applications that store IP addresses within the application, and communicate these addresses to a remote host. Such applications would not function properly under our architecture. They are readily identifiable, however, as another currently widely-deployed technology also breaks such applications: Network Address Translators (NATs). While the wisdom of Network Address Translation is a hotly debated topic, there is little chance it will disappear any time soon. Hence most applications designed today take care not to transmit addresses as part of the application-layer communication, and therefore will also work in our architecture. In fact, one can make the case that such applications are broken, since IP addresses are only identifiers of attachment points, not hosts.

Another, larger class of applications cache the results of *gethostbyname()*, and may not perform further hostname resolution.[8] Furthermore, DNS resolvers themselves cache hostname bindings as discussed in Section 3. Unfortunately many older name servers enforce a local TTL minimum, often set to five minutes. Since newer versions of popular name servers adhere to the TTL specified in the returned resource record, this problem should disappear as upgrades are made.

7.2 Proxies and NATs

Proxies actually help the deployment of our scheme, as we only need to modify the proxy itself, and all communications through the proxy will support mobility. Similarly, NATs can also provide transparent support without remote system modification. In fact, a NAT doesn't even need a modified TCP stack. It need only snoop on TCP SYNs (which it does anyway), note the presence of a Migrate-Permitted option, and snoop for the SYN/ACK (which it does anyway). If the SYN/ACK does not contain a Migrate-Permitted option, the NAT can support connection migration internal to its network by inserting a corresponding Migrate-Permitted option, and continuing to snoop the flow looking for any Migrate SYNs. It need only fabricate a corresponding SYN/ACK and update its address-to-port mappings, without passing anything to the end host. Further, by avoiding any explicit addressing in migrate requests, the Migrate options function properly though legacy NATs, and even allow a mobile host to move between NATs, as connections may change not only address but port as well.

7.3 Non-transactional UDP applications

Many UDP applications are transactional in nature. UDP is, by definition, a datagram protocol, and an inopportune change of IP address is only one of many reasons for an unsuccessful UDP transaction. The transaction will need to be retried, although a new hostname binding should be obtained first.

There exists at least one glaring exception to this rule. The Network File System protocol (NFS) represents one of the most prevalent UDP applications in use today and uses IP addresses in its mount points.[9] We believe, given the characteristics of network links likely to be encountered by mobile hosts, it is likely that NFS-over-TCP is a better choice than UDP. Otherwise, a mobile host would need to dismount and re-mount NFS filesystems upon reconnection.

8 Conclusion

This paper presents an end-to-end architecture for Internet host mobility that makes no changes to the underlying IP communication substrate. It uses secure updates to the DNS upon an address change to allow Internet hosts to locate a mobile host, and a set of connection migration options to securely and efficiently negotiate a change in the IP address of a peer without breaking the end-to-end connection. We have implemented this architecture in the Linux operating system and are encouraged by the ease with which mobility can be achieved without any router support, the flexibility to mobile hosts provided by it, and performance comparable to Mobile IP with route optimization.

Our architecture allows end systems to choose a mobility mode best suited to their needs. Routing paths are efficient with no triangle routing, and any connection involving the mobile host shares fate only with the communicating peer and not with any other entity like a home agent. When a mobile host is in a foreign network and communicating with another host, the disruption in connectivity caused by a sudden IP address change is proportional to the round-trip time of the connection. When a mobile host accepts no passive connections, the protocol does not require even the DNS update notification, and seamless connectivity across host mobility is achieved using completely end-to-end machinery.

The security of our approach is based on a combination of the well-documented secure DNS update protocol in conjunction with a new secure connection migration mechanism. Our architecture and implementation function across a variety of other components of the Internet architecture, including firewalls, NATs, proxies, IPsec, and IPv6. We believe that our architecture scales well even when most Internet hosts become mobile because lookups and updates are distributed across administratively-delegated, replicated DNS servers.

We note that our connection migration scheme, the MIGRATE_WAIT state in particular, avoids address assignment race conditions, but does *not* support host disconnectivity. Hence, as with Mobile IP and other mobility schemes, TCP connections may be lost if the mobile host's relocation is accompanied by a prolonged period of disconnectivity. We are hopeful our end-to-end approach may be extended to support general host disconnectivity as well.

Acknowledgements

We thank John Ankcorn, Frans Kaashoek, Eddie Kohler, Robert Morris, Srinivasan Seshan, Tim Sheppard, and Karen Sollins for helpful comments on earlier drafts of this paper. We are indebted to David Andersen, who helped improve the security of our initial Migrate scheme, and David Mazieres, who suggested we use Elliptic Curve Diffie-Hellman key exchange for additional key strength.

[8] Some popular Web browsers display this behavior.

[9] We note that most other advanced file systems, such as Coda [22] and newer versions of NFS use TCP, which gives good congestion control and reliability behavior.

RED HAT 1010
RED HAT V. CAS

References

[1] ADJIE-WINOTO, W., SCHWARTZ, E., BALAKRISHNAN, H., AND LILLEY, J. The design and implementation of an intentional naming system. In *Proc. ACM SOSP '99* (Dec. 1999), pp. 186–201.

[2] AKAMAI TECHNOLOGIES, INC. `http://www.akamai.com`.

[3] AMERICAN NATIONAL STANDARDS INSTITUTE. Public key cryptography for the financial service industry: The elliptic curve digital signature algorithm. ANSI X9.62 - 1998, Jan. 1999.

[4] ATKINSON, R. Security architecture for the internet protocol. RFC 1825, IETF, Aug. 1995.

[5] BALAKRISHNAN, H., SESHAN, S., AND KATZ, R. H. Improving reliable transport and handoff performance in cellular wireless networks. *ACM Wireless Networks 1*, 4 (Dec. 1995), 469–481.

[6] CACERES, R., AND IFTODE, L. Improving the performance of reliable transport protocols in mobile computing environments. *IEEE JSAC 13*, 5 (June 1995).

[7] DROMS, R. Dynamic Host Configuration Protocol. RFC 2131, IETF, Mar. 1997.

[8] EASTLAKE, 3RD, D. E. Secure domain name system dynamic update. RFC 2137, IETF, Apr. 1997.

[9] FERGUSON, P., AND SENIE, D. Network ingress filtering: Defeating denial of service attacks which employ IP source address spoofing. RFC 2267, IETF, Jan. 1998.

[10] GUPTA, S., AND REDDY, A. L. N. A client oriented, IP level redirection mechanism. In *Proc. IEEE Infocom '99* (Mar. 1999).

[11] HUITEMA, C. Multi-homed TCP. Internet Draft, IETF, May 1995. (expired).

[12] IEEE. Wireless medium access control (MAC) and physical layer (PHY) specifications. Standard 802.11, 1999.

[13] JACOBSON, V. Congestion avoidance and control. In *Proc. ACM SIGCOMM '88* (Aug. 1988), pp. 314–329.

[14] JACOBSON, V., BRADEN, R., AND BORMAN, D. TCP extensions for high performance. RFC 1323, IETF, May 1992.

[15] JOSEPH, A. D., TAUBER, J. A., AND KAASHOEK, M. F. Mobile computing with the rover toolkit. *IEEE Trans. on Computers 46*, 3 (Mar. 1997), 337–352.

[16] KARN, P. Qualcomm white paper on mobility and IP addressing. `http://people.qualcomm.com/karn/papers/mobility.html`. Feb. 1997.

[17] LENSTRA, A. K., AND VERHEUL, E. R. Selecting cryptographic key sizes. `http://www.cryptosavvy.com`. Nov. 1999.

[18] MALTZ, D., AND BHAGWAT, P. MSOCKS: An architecture for transport layer mobility. In *Proc. IEEE Infocom '98* (Mar. 1998).

[19] MATHIS, M., MAHDAVI, J., FLOYD, S., AND ROMANOW, A. TCP selective acknowledgment options. RFC 2018. IETF, Oct. 1996.

[20] MOCKAPETRIS, P. V., AND DUNLAP, K. Development of the domain name system. In *Proc. ACM SIGCOMM '88* (Aug. 1988), pp. 123–133.

[21] MORRIS, R. T. A weakness in the 4.2BSD UNIX TCP/IP software. Computing science technical report 117. AT&T Bell Laboratories, Murray Hill, New Jersey, Feb. 1985.

[22] MUMMERT, L. B., EBLING, M. R., AND SATYANARAYANAN, M. Exploiting weak connectivity for mobile file access. In *Proc. ACM SOSP '95* (Dec. 1995), pp. 143–155.

[23] MYSORE, J., AND BHARGHAVAN, V. A new multicasting-based architecture for internet host mobility. In *Proc. ACM/IEEE Mobicom '97* (Sept. 1997), pp. 161–172.

[24] NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY. The Secure Hash Algorithm (SHA-1). NIST FIPS PUB 180-1, U.S. Department of Commerce. Apr. 1995.

[25] NOBLE, B. D., SATYANARAYANAN, M., NARAYANAN, D., TILTON, J. E., FLINN, J., AND WALKER, K. R. Agile application-aware adaptation for mobility. In *Proc. ACM SOSP '97* (Oct. 1997), pp. 276–287.

[26] PERKINS, C. E., AND CALHOUN, P. R. Mobile IP challenge/response extensions. Internet Draft, IETF, Feb. 2000. `draft-ietf-mobileip-challenge-09.txt` (work in progress).

[27] PERKINS, C. E., AND JOHNSON, D. B. Mobility support in IPv6. In *Proc. ACM/IEEE Mobicom '96* (Nov. 1996), pp. 27–37.

[28] PERKINS, C. E., AND JOHNSON, D. B. Route optimization in mobile IP. Internet Draft, IETF, Feb. 2000. `draft-ietf-mobileip-optim-09.txt` (work in progress).

[29] PERKINS, ED., C. E. IP mobility support. RFC 2002, IETF, Oct. 1996.

[30] POLLARD, J. Monte carlo methods for index computation mod p. *Mathematics of Computation 32* (1978), 918–924.

[31] POSTEL, ED., J. Transmission Control Protocol. RFC 793, IETF, Sept. 1981.

[32] SALTZER, J. H., REED, D. P., AND CLARK, D. D. End-to-end arguments in system design. *ACM TOCS 2*, 4 (Nov. 1984), 277–288.

[33] STEVENS, W. R. *TCP/IP Illustrated, Volume 1: The Protocols*. Addison Wesley, Reading, Massachusetts, 1994.

[34] THOMSON, S., AND NARTEN, T. IPv6 stateless address autoconfiguration. RFC 2462. IETF, Dec. 1998.

[35] VIXIE, P., THOMSON, S., REKHTER, Y., AND BOUND, J. Dynamic updates in the domain name system (DNS UPDATE). RFC 2136, IETF, Apr. 1997.

[36] ZUCCHERATO, R., AND ADAMS, C. Using elliptic curve Diffie-Hellman in the SPKM GSS-API. Internet Draft, IETF, Aug. 1999. `draft-ietf-cat-ecdh-spkm-00.txt` (work in progress).

RED HAT 1010
RED HAT V. CAS

A New Multicasting-based Architecture for Internet Host Mobility

Jayanth Mysore Vaduvur Bharghavan
Coordinated Sciences Laboratory
University of Illinois at Urbana-Champaign
{babu, bharghav}@timely.crhc.uiuc.edu

Abstract

IP multicasting provides a mechanism for location independent addressing and packet delivery to a group of hosts that belong to a multicast group. It also provides efficient mechanisms for hosts to join and leave multicast groups. The problem of supporting macro and micro level host mobility in the Internet involves similar issues of location independent addressing, address translation, packet forwarding and location management of mobile hosts.

In order to exploit the commonality of goals in the two problems, we have investigated a new approach for supporting host mobility using IP multicasting as the sole mechanism for addressing and routing packets to mobile hosts. Mobile hosts are identified by unique multicast IP addresses, and use the IP multicasting infrastructure to route packets destined for mobile hosts. We have implemented a mobile computing environment which uses this routing approach, and our performance analysis shows that it is very effective in supporting seamless mobility for both handoffs and interface changes within and across networks.

We discuss a number of issues which need to be resolved in order to make our approach viable and practical for supporting host mobility in the Internet.

1 Introduction

In the Internet addressing format, the unicast IP address of a host identifies the host uniquely and also identifies the network to which it is attached. Thus, a unicast IP address plays the dual role of both end-point identification and location identification. While the former role is used by TCP for connection management (a TCP connection is uniquely identified by the 4-tuple <src addr, src port, dst addr, dst port>), the latter role is used by IP for routing packets to the host. Thus, when a mobile host changes it's point of attachment from one network to another, it's unicast address must be modified in order to reflect this change. The evolution of TCP/IP networks did not consider the problem of mobility until recently. The TCP connection identifier is set up during connection establishment, and is not expected to change during the lifetime of the connection. Special mechanisms are therefore needed to support mobility while still allowing ongoing TCP connections to transmit and receive packets transparently. While simple changes in the TCP protocol which allow the connection identifier to change dynamically during a connection lifetime can solve the mobility problem, this will involve changing the protocol stack on every host and notifying all connection end-points when a mobile host moves. Thus, Mobile IP and its variants seek to provide a network-level solution which enables static hosts to remain oblivious to the dynamic address changes in the mobile hosts. In Mobile IP, a mobile host has two addresses: an invariant *home address*, which is used for end-point identification, and a dynamic *care-of address*, which is used for location identification. While TCP uses the home address for connection identifiers, IP uses the care-of address for routing. Mobile IP maps the home address to the care-of address, and hence provides location independent addressing and routing mechanisms for mobile hosts.

Interestingly, the problem of *multicasting* faces the identical challenge of providing location independent addressing and routing support, though in a different context. In IP multicasting, hosts spread across different networks can join a multicast group, and can receive packets that are sent to the group. A message sent to a multicast group does not embed the location of the destination hosts. Hence, location independent addressing and routing mechanisms are necessary to support multicasting.

The similarity in the fundamental nature of the two problems has resulted in the development of almost identical architectures for supporting the two services. The architectures are compared in Table 1. The commonality of the issues that need to be addressed in mobility support and multicasting caused us to explore the possibility of using the evolving Internet multicasting infrastructure in order to support host mobility. In particular, our goal is *to use multicasting as the sole mechanism to provide addressing and routing services for mobile hosts.* We assign a unique multicast address to each mobile host - consequently, all packets from a correspondent host to a mobile host are treated as multicast packets, while packets from a mobile host to a static host are treated as standard unicast packets. *The use of multicast addresses to identify mobile hosts enables us to support mobility using the multicasting infrastructure with minimal or no changes to the*

RED HAT 1010
RED HAT V. CAS

backbone networks. This is a major departure from contemporary approaches for supporting mobility [3], which typically use multicasting only as a support mechanism for reducing packet loss during handoff.

In this paper, we first show that the fundamental issues faced by mobile host support are a sub-set of the issues faced by any network level multicasting approach. We then present an architecture that uses multicasting to support mobility. We describe our implementation testbed and present performance results which show that our approach performs well for both handoffs and interface changes within and across subnetworks. This leads us to conclude that our approach could provide a viable alternative to current state-of-the-art solutions. Finally, we discuss several key issues that need to be resolved before our approach can be deployed in a widespread manner in the Internet. Our experiences indicate that the multicasting technology is still not mature enough to support ubiquitous host mobility in the Internet; however, the fundamental approach of using multicasting to support host mobility not only solves the problem of supporting mobility within and across networks efficiently, but also has several desirable properties in terms of resource reservation in future Integrated Services Networks, reducing handoff dropping, and advance registration/routing.

It is important to note that we are *not* claiming that the current IP multicasting infrastructure is sufficient to adequately support mobility in the Internet. Rather, the goal of this paper is to show that any viable multicasting infrastructure for the next generation Internet will need to address all the issues that are germane to the problem of supporting host mobility in the Internet (besides other issues that are unique to multicasting). Hence, we suggest using a multicasting infrastructure to solve the problem of mobility support essentially for free.

The rest of the paper is organized as follows. Section 2 presents the equivalence between the multicasting and mobility problems. Section 3 describes the multicasting architecture to support mobility. Section 4 describes our testbed implementation and presents the performance results. Section 5 discusses key issues which need to be resolved before an IP multicasting infrastructure can be used in practice to support Internet host mobility. Section 6 compares our approach with related work, and Section 7 concludes the paper.

2 Equivalence of Multicasting and Mobility Support

2.1 The IP-Multicasting Approach

The purpose of multicasting is to enable transmitters to send packets to multiple hosts in an efficient manner. Multicasting thus provides for a common communication channel with multiple receivers. In order to provide this service, the following issues need to be addressed.

1. Multicast IP addresses should identify location independent communication channels rather than location specific host addresses.
2. Packets with multicast destination addresses should be forwarded to all hosts in the multicast group irrespective of their location.
3. Mechanisms for locating, updating, and managing multicast groups, addresses and membership need to be developed.
4. Hosts should have the ability to dynamically join and leave multicast groups.

The IP multicasting approach addresses the above issues by assigning a multicast address to a group of receivers. Senders simply use the multicast address to send packets to the group. Specialized multicast routers are created in every subnet that supports multicast recipients. A group membership protocol called IGMP is used for routers to learn the existence of members of multicast groups in their directly attached subnets. Multicast routers establish a virtual network among themselves, on top of the underlying internetwork. 'Links' in the virtual network between any two multicast routers are tunnels between the two routers which may span multiple unicast links. Multicast routers exchange information between themselves in order to maintain and propagate membership information of different multicast groups. When a host addresses a packet to the multicast group, the multicast router on its local subnet picks up the packet, and assumes the burden of delivering the packet to all members of the multicast group. Several candidate multicast routing protocols, such as DVMRP [23], MOSPF [15], PIM [8], CBT [2], etc. are available to perform the routing among the multicast routers. Eventually, every multicast router which has at least one member receives the packet, and locally multicasts it in the subnet. While a detailed discussion of the candidate multicast routing algorithms is beyond the scope of this paper, the common distinguishing feature of all these algorithms is that they use the multicast router virtual network in order to tunnel packets to the destinations. End hosts receive multicast packets as a result of a directory-based packet forwarding mechanism rather than the use of location-specific unicast destination addresses. Figure 1.a illustrates the routing mechanism.

2.2 The Mobility Support Approach

In order to support host mobility, the IP address needs to be location independent and invariant. However, for IP to route packets, the host address needs to reflect the network address of the current point of attachment. In order to reconcile these two conflicting requirements, most contemporary approaches provide a two-level addressing scheme. While the different approaches for providing mobility support are outlined in Section 6, the generic paradigm is to use packet forwarding and address translation *agents* in order to tunnel packets to the current location of the mobile host. Each subnet that supports mobility provides an *agent* (or at least a mechanism to provide a location specific temporary address). For a mobile host, the agent in its home network (identified by its invariant home address) is its home agent and the agent

RED HAT 1010
RED HAT V. CAS

Functionality	Mobile Host Support	Multicasting
Registration	A mobile host that enters a new network must register with the local agent.	A host that wishes to receive multicast datagrams must register with the local multicast router.
Connectivity	Connectivity to the rest of the internet is provided by the foreign agent.	A multicast router provides connectivity to the rest of the virtual multicast network to a host.
Data forwarding	The foreign agents forward datagrams to the mobile host.	The multicast router multicasts datagrams to the members in its subnet.
Address translation	Home agents translate the home address to the mobile host's care-of address.	Multicast messages need no address translation
Routing	The home agent tunnels messages destined to the mobile host to the current foreign agent.	Messages sent to a destination multicast address are forwarded to multicast routers with members using a multicast routing protocol.

Table 1: Comparison of Multicasting and Host Mobility Architectures

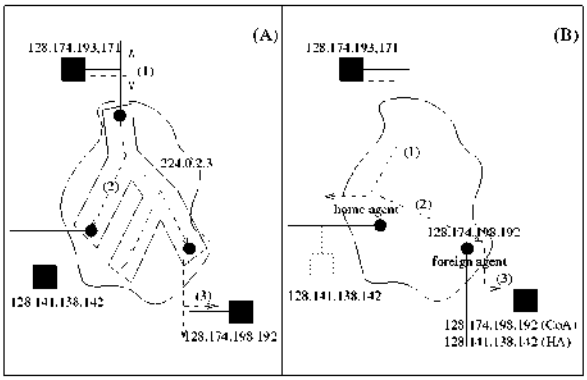


Figure 1: Figure (a) shows the multicast model. In Step (1), a host transmits a multicast packet. In Step (2), the multicast router tunnels the multicast packet along the multicast distribution tree. In Step (3), each destination multicast router decapsulates the tunneled packet and multicast it locally in its subnet. Figure (b) shows the mobility model. In Step (1), the correspondent host transmits the packet to the home address of the mobile host. In Step (2), the home agent tunnels the packet to the foreign agent. In Step (3), the foreign agent decapsulates the tunneled packet and forwards it to the mobile host.

in its current network (identified by it's care-of address) is its foreign agent.

When a mobile host moves to a different network, it registers the care-of address (foreign agent) with the home agent. Hosts on the Internet are aware only of the home address of the mobile host. When the home agent sees a packet addressed to a mobile host that is registered with a foreign agent, it tunnels the packet to the foreign agent, which then forwards the packet to the mobile host. Uplink transmissions from the mobile host to the correspondent host may be directly forwarded to it using standard IP mechanisms. In the above architecture, the agents form a virtual network among themselves, essentially providing a similar type of location independent service to mobile hosts, like the multicast routers provide to the members of a multicast group.

The basic motivation of our work is the observation that the mobility problem is subsumed in the multicasting problem. Both problems require the following key issues to be addressed: *providing the abstraction of location independent addressing*, *packet forwarding*, and *location management*. Several contemporary approaches for providing mobility support have been proposed. Bhagwat et al [3] provide a survey of some of these approaches, and compare them based on their solutions to the above issues. Among these approaches, Mobile IP is emerging as the standard for mobility support in IPv4. In Mobile IP, address translation is performed by the home agent (mapping home address to the care-of address); packet forwarding is done by the home agent and foreign agent; location management is done by the home agent.

It has been noted in both [17] and [19], that while Mobile IP (and IPv6) is designed to address the macro-level mobility problem such as supporting host mobility over wide area networks, it does not address micro-level mobility issues such as packet loss and delay due to handoffs. We address the issues at the two levels in an integrated approach, by using the emerging multicasting infra-structure to provide support for host mobility in the Internet. In our approach, we *assign unique Class-D addresses to each mobile host. Messages sent to a mobile host will therefore be treated like multicast messages, and routed through a network of multicast routers to the host.*

In the rest of this paper, we refer to our approach as MSM-IP, which stands for Mobility Support using Multicasting in IP.

2.3 MSM-IP versus Mobile IP

While the problems of supporting multicasting and mobility are very similar, there are several differences in the state-of-the-art approaches for IP multicasting and Mobile IP. We first contrast IP multicasting with Mobile IP, and then compare it to IPv6.

The MSM-IP approach for supporting mobility differs from Mobile IP is five important ways: addressing, packet

RED HAT 1010
RED HAT V. CAS

forwarding, location management, service disruption, and advance reservation/routing.

Addressing: The IP multicasting approach to supporting mobility issues a single invariant address to a mobile host. This eliminates the need for explicit address translation as in Mobile IP. On the other hand, a mobile host does not have a home network. In particular, the behavior of the mobile host and the office desktop machine of a mobile user will be different (for example, the mobile host will not receive broadcast packets on the user's office network, it belongs to a different administrative domain, etc). An obvious limitation of this approach is the size of the Class D address space in IPv4.

Packet Forwarding: In Mobile IP, since only the home agent is aware of the care-of address of the mobile host, all packets from a correspondent host to a mobile host suffer from triangular routing. In the MSM-IP approach, the multicast router on the local subnet of the correspondent host will pick up the packet and forward it directly to the multicast router of the mobile host's subnet. Thus, packets will not suffer triangular routing. On the other hand, Mobile IP does not require a mobility-aware agent on a subnet that does not support mobile hosts, while IP multicasting requires a multicast router on every subnet that supports either senders or receivers of multicast packets.

Location Management: In Mobile IP, correspondent hosts use the home address of a mobile host. Thus, it is sufficient for a mobile host to notify its home agent of its current care-of address. In MSM-IP, there is no notion of a home agent. When any multicast router wishes to forward a packet to a mobile host, it needs to locate the host using a distributed directory algorithm. Development of a scalable location management protocol (or session discovery protocol for standard multicast transmissions) is still ongoing work in IP multicasting research. We describe our approach to location management in Section 3.

Service Disruption: As mentioned in the IETF drafts, Mobile IP attempts to solve the problem of 'macro-level' mobility rather than 'micro-level' mobility [19]. Thus, critical issues such as minimizing packet loss for on-going TCP connections during mobility are not stressed. In the MSM-IP approach for supporting mobility, we are concerned with both macro-level and micro-level mobility. Our claim is that MSM-IP (in effective sparse mode operation [8]) could perform as well as Mobile IP for macro level mobility, while significantly reducing the disruption in service during handoffs. The main reason for this claim is the following: in Mobile IP, response to a mobile host's change in location occurs only after the home agent (correspondent host in IPv6) is made aware of the change. However, in MSM-IP, all joins and prunes are receiver-initiated and terminate at the earliest branch point. Thus, response to a mobile host's change in location happens as close to the mobile host as possible, which significantly reduces the latency of response and disruption of service for long haul connections.

Advance Reservation/Routing: Several contemporary approaches have used multicasting in a more restricted form than we propose, for the purpose of reducing or eliminating packet loss during handoffs. Before a mobile host performs a hand-off, it may request the multicasting router in the new network to join the multicast group given by its address, in advance. Upon handoff, the mobile host will continue to receive packets uninterrupted. The overhead for such advance joins is expected to be tolerable, particularly in the common case of handoff between adjacent cells, since the old and the new base stations will be either in the same subnet or neighboring subnets. Similar to an advance join, it is also possible for the multicast router to initiate advance resource reservation, as described in [14].

MSM-IP Vs IPv6: IPv6 has the luxury of being able to change the IP behavior at correspondent hosts as well as mobile hosts. A current approach for supporting mobility in IPv6 is to make correspondent hosts aware of the care-of address of a mobile host as it moves. Thus, triangular routing is avoided. However, for long haul connections, service will still be disrupted upon a handoff until the correspondent host is notified about the new care-of address of the mobile host. The comparison with IPv6 is not the focus of this work, since our goal is to be as minimalist as possible in the changes we make at the backbone, and still provide mobility support and reduced packet loss/service disruption upon handoff. In the next section, we describe our IP multicasting architecture for the support of mobility.

3 Architecture

The starting premise of this work is that host mobility can be supported without making any special changes to an Internet multicasting architecture, though the state-of-the-art of IP multicasting is still inadequate for this purpose. With this in mind, we present an architecture for supporting mobility using state-of-the-art IP multicasting in this section, describe our experience with such an architecture in Section 4, and raise several issues that need to be addressed before the proposed architecture can be used in practice in Section 5.

The IP multicasting architecture consists of three types of hosts, *transmitters*, *multicast routers*, and *receivers*. A transmitter need not be a member of a multicast group to transmit to the group. Every subnet that supports either the transmission or reception of multicast packets must have a multicast router. The multicast router is typically assumed to be colocated with the base station, though it may be any host on the same subnet as the base station (for example, we can have multiple Wavelan access points attached to the same subnet served by a single multicast router). The multicast routers form a virtual network among themselves, in order to exchange group membership information, perform location/session discovery, and

RED HAT 1010
RED HAT V. CAS

forward packets.

When a receiver wants to join a multicast group, it initiates a join by sending an IGMP registration message to the multicast router in its subnet. Each multicast router maintains a distributed directory of group membership information. When a transmitter wants to send a packet to a multicast group, it sets the IP destination address to the group address and sends the packet as a local multicast. When it transmits the packet, the multicast router in its subnet picks up the packet. It is now the responsibility of this multicast router to propagate the packet to the recipients in the multicast group. Multicast routing within the network of multicast routers is performed in several ways. For example, using a distance vector protocol such as DVMRP [23], the goal is to construct a multicast distribution tree for each group based on Truncated Reverse Path Broadcasting [6]. While DVMRP works well in the *dense mode*, wherein several members of the multicast group are geographically close together, supporting mobility typically will involve multicasting in the *sparse mode* with few receivers (in this case, one, in the absence of advance registration). Several protocols such as PIM-SM [8] and CBT [2] address the issues of sparse mode operation. In this section, we present a generic architecture for IP multicasting, that can be used to support host mobility effectively. The routing protocol used in this architecture is beyond the scope of this paper, though we expect a protocol such as PIM-SM to work well so long as the Rendezvous points are chosen carefully (see Section 3.4 on location management).

We now describe the seven key components of our IP multicasting based host mobility support architecture: (a) addressing, (b) tunneling architecture, (c) join and prune mechanisms, (d) location management, (e) handoff and advance registration, (f) resource reservation, and (g) supporting and interacting protocols such as ARP, IGMP, ICMP, RARP, TCP and UDP.

3.1 Addressing

Static hosts are assigned standard unicast IP addresses. Mobile hosts are assigned unique multicast IP addresses. Thus, packets from a mobile host to a static host will be routed by standard unicast mechanisms while packets from a static host to a mobile host will be routed by the multicasting infrastructure. There are two key issues in this approach: (a) some routers will not forward packets that originate from hosts with addresses that do not correspond to the local network- such routers will drop packets from the mobile hosts, and (b) mobile hosts cannot addresses each other directly if the multicast routers use reverse path forwarding (RPF) (since RPF assumes that the source address is a unicast IP address in order for multicast routers to determine the shortest path to the source). Section 5 discusses these issues and provides initial solutions.

3.2 Tunneling Architecture

Each subnet that supports multicast transmission and reception has a multicast router. Each multicast router is aware of the other routers in its neighborhood. Multicast routers exchange routing information on-demand, and build multicast packet distribution trees for each multicast group. The mobile host will initiate a group membership registration with the multicast router in its subnet, and potentially some other neighboring multicast routers (advance registration). Thus, each multicast group corresponding to a mobile host's address has a few clustered recipients. Transmitters to this group may be anywhere in the network. Using sparse mode multicast routing algorithms [8, 2], it is possible to create a multicast distribution tree for the multicast group. Each node in the tree has information about which nodes are downstream, i.e. when it receives a packet for a multicast group, to which multicast routers it should forward these packets. The design of algorithms for building efficient multicast distribution trees is still ongoing research. It should be noted that any algorithm that works well for sparse mode multicasting will work equally well for supporting mobility.

3.3 Join and Prune Mechanisms

In order to provide support for hosts to dynamically join and leave multicast groups, the IP multicasting architecture uses a soft-state based approach for membership registration. Hosts join a multicast group by registering with the multicast router through the IGMP Registration message. Periodically, hosts also send a registration refresh message to the router (when a host sees an IGMP registration message for the multicast group it belongs to, it resets its timer; thus, only one host on a subnet needs to send the refresh during a timeout period). When a multicast router (which is not a leaf of the multicast distribution tree) receives prune messages from all its children, and does not have any hosts in the multicast group in its own subnet (either it never had members in its subnet or their registration timed out), it prunes itself from the multicast distribution tree by sending a prune message to its parent(s).

Receiver-initiated joins and prunes, and the use soft-state to time out registration work very well to support host mobility. When a host enters a new domain, it initiates a join for its (multicast) address. The multicast router in its subnet then joins the multicast distribution tree. In the common case of handoff, we anticipate that either the multicast router would already have joined the multicast group through advance registration, or at least is a neighbor of the previous multicast router serving the mobile host. Thus, most receiver-initiated joins terminate at the lowest levels of the distribution tree hierarchy. Once a mobile host enters a new cell and registers with its new multicast router, it ceases to send refreshes to the old multicast router. Eventually, this router will time out its registration and initiate a prune. As in the case of joins, we expect most prunes to terminate at the lowest levels of the distribution tree hierarchy. Thus, handoff using the IP multicast architecture is very efficient. As shown in Section 4, and also in related work [20], using multicasting for handoff can reduce or eliminate transient packet loss.

RED HAT 1010
RED HAT V. CAS

3.4 Location Management

A critical issue that needs to be addressed if IP multicasting is to be deployed in a widespread manner is multicast session discovery. The first problem relates to discovering what sessions are ongoing in the Internet at any time such that hosts can join these sessions - this is a resource discovery problem. The second problem relates to discovering the nearest multicast router to join in order to receive multicast packets addressed to a group - this is a location update problem. In the current IP multicasting architecture, these problems are not adequately handled. In particular, the resource discovery issue is handled by means of distributed session directories while the location management problem is handled by flooding the network to find the nearest multicast router in the worst case [23]. In order to support mobility in the IP multicasting architecture, the location update protocol needs to be sophisticated. We have designed a location directory protocol which is similar to the Mobile IP approach for location discovery. We divide the space of multicast addresses into *hierarchical virtual domains*, where each domain is served by a location server. The location server that serves the multicast address of a mobile host contains the location of a multicast router which 'knows' how to reach the mobile host (described in detail below). Route discovery to a mobile host in our environment is thus a 5 step process, as shown in Figure 2: (i) the correspondent host (CH) sends a multicast packet addressed to the mobile host (MH), (ii) the multicast router (M1) in the transmitter's subnet picks up the packet, (iii) M1 discovers the location server (LS) of the mobile host based on its multicast address (through a location discovery protocol that maps multicast addresses to location servers), (iv) M1 contacts LS and discovers the current multicast router entry (M2) corresponding to MH, (v) M1 communicates with M2 in order to join the multicast distribution tree.

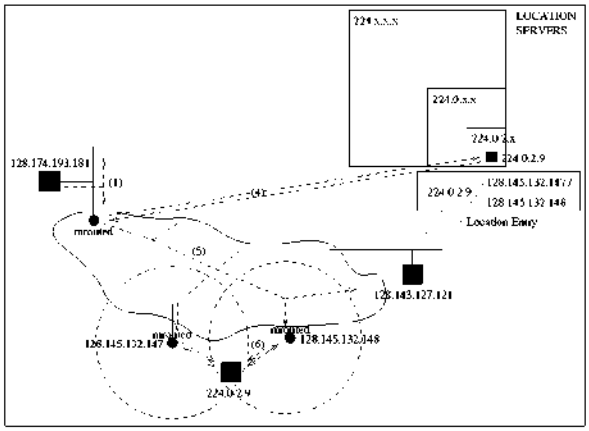


Figure 2: The location entry shows the contents at a location server corresponding to a mobile host. In Step (1), the correspondent host transmits a packet to the mobile host. In Step (2), the multicast router picks up the packet. In Step (3), it discovers the location server corresponding to the mobile host. In Step (4), it contacts the location server and retrieves the address of a multicast router which can forward packets to the mobile host. In Step (5), the transmission multicast router joins the multicast distribution tree and transmits the packet. In Step (6), each receiving multicast router decapsulates the packet and forwards it to the mobile host.

Note, that while the architecture specifies the overall scheme for location discovery and updates, it does not specify the specific algorithms for either location server discovery in step (iii), the multicast distribution tree in step (v), or which multicast router is associated with a mobile host in the location server.

In a PIM-SM approach, for example, the rendezvous point would be identified as the multicast router corresponding to the mobile host. In our approach, we identify the current (or a recent) multicast router of the mobile host. When a mobile host moves, it periodically refreshes the multicast router in its location server. Note, that even after a mobile host moves out of its domain, the old multicast router will retain membership registration for a timeout period. So long as the mobile host sends location updates messages to its location server at a frequency greater than twice each timeout period, new transmitters which are not aware about the location of a mobile host can reach the host by looking at the multicast router information from its location server.

We believe that our location management architecture is a generic approach which can be integrated with several emerging approaches for IP multicast routing.

3.5 Handoff and Advance Registration

As mentioned above, the IP multicasting support for receiver initiated join/prune and soft-state based timeout provides effective mechanisms for initiating handoffs. One of the key advantages of multicasting is the ability to use advance registration with the new cell (or multicast router in the new domain) in order to minimize packet loss during handoff. We have used this mechanism in concert with our resource reservation approach in order to provide advance resource reservation in mobile computing environments [14].

3.6 Resource Reservation

One of the major advantages of using multicasting as the sole mechanism for mobility support is our ability to use RSVP for resource reservation and advance reservation. As noted in [24], one of the goals of RSVP is to support resource reservations in the presence of dynamically joining and leaving receivers. While recent approaches have tried to tune RSVP in order to run effectively in mobile computing environments (given that they take the Mobile IP approach for supporting mobility and still seek to perform advance reservations using RSVP), we can run RSVP without any modifications in our architecture for supporting mobility.

RED HAT 1010
RED HAT V. CAS

3.7 Interacting Protocols

Conceptually, the IP multicasting architecture can support host mobility without any changes to the existing infrastructure by simply assigning a unique multicast IP address to a mobile host. However, a number of interacting protocols make the practical implementation of such an architecture difficult without making any changes to any component either at the mobile host or the multicast routers. The following are the key protocols which interact with the IP routing protocol: ARP, RARP, ICMP, IGMP, TCP, and UDP. Among these, ARP replies sent to multicast IP addresses are not processed by the ARP module on the portable. ICMP messages cannot be routed to hosts with multicast addresses. TCP needs to change at the mobile host in order to accept packets with a multicast destination address. Unless TCP checks for a multicast source address (which some kernels do [21]), TCP at the static host does not need to change. UDP does not need to change in order to operate using multicast addresses, though a higher layer protocol needs to address the issue of lost packets and duplicates in UDP packets upon handoff or interface change. Details of the problems and the initial solutions are provided in Section 6.

The change in TCP and IGMP at the mobile host are trivial and no new functionality has to be added[1]. While the above set of problems seems like a large list of incompatible protocols, note that the key problems stem from a single fact - they require a location-specific host IP address for network management or network layer processing. In other words, while the use of a multicast address is very effective in order to provide location independence and support mobility, several IP level functionalities in fact require the location specific nature of unicast IP addressing to be explicitly exposed (e.g. ARP and ICMP). There are two ways to solve the above problems - (a) provide additional functionality in multicast routers, or (b) use DHCP in order to obtain a temporary unicast address at the mobile host for the specific purpose of handling network level management. We adopt the latter approach because it is much simpler, and provides a natural solution to the routing problems - use location independent addresses to support mobility, and use location dependent addresses to support network management. Thus, with minimal changes in the supporting protocols and no change to the basic IP multicasting architecture, we can support host mobility effectively, and provide several desirable features such as advance registration, resource reservation and no transient packet loss during handoff.

4 Implementation and performance results

The MSM-IP architecture has been implemented and operational in our laboratory testbed environment for about six months now. The testbed backbone consists of three switched Ethernet subnets, which are served by gateways *durga*, *atri* and *radha* respectively. *durga* is connected to *atri* and *radha*, which are not connected directly to each other. We have two Wavelan wireless subnets, whose gateways are attached to the backbone subnets of *atri* and *radha* respectively. All the correspondent hosts used for testing are P6-200 Gateway Pentium Pros, and all the mobile hosts used for testing are P6-120 TI laptops. All the hosts run Linux 2.0.30. *durga*, *atri* and *radha* run the `mrouted` 3.81 multicast router. In all our experiments, the correspondent hosts are located in the subnet of *durga*, while the mobile hosts move between the wireless subnets of *atri* and *radha*. All mobile hosts have both Ethernet and Wavelan interfaces. Note that a handoff or interface switch between subnets is accompanied with a change in path from *durga* (i.e. at least two hops). Additionally, we have instrumented *durga* to optionally add a 100ms delay to all packets passing through it in order to simulate network latency.

Each mobile host runs a `registration` module which can register with the current `mrouted` and the `mrouted` serving the next cell via the backbone network (for advance registration), and send periodic refresh messages to keep alive the registrations. The applications we run in the testbed include WWW Browsers, ftp, Mpeg video over the network, editors, and file system access as well as special workloads generated for the purpose of quantifying the performance of our approach. Since the standard TCP does not work when one of the end-points is a Class D address, we made simple changes to the kernel to disable these checks for the purposes of doing performance measurements. We also instrumented TCP to record round trip time measurements. Our TCP-based applications thus use this modified TCP protocol, while while our UDP based applications use the standard UDP protocol.

We have measured the performance of MSM-IP with variations in offered workload, artificial delay over the network, applications, types of handoff, and parameters of measurement. In this paper, we present a subset of these measurements. In terms of offered load, we used UDP flows with offered throughputs of 10Kbps, 100Kbps, 512 Kbps and 1Mbps, and TCP flows at peak traffic. In terms of packet sizes, we used 512 bytes and 1 KB packets for the user generated workloads, and also used the application programs (WWW browser, ftp, Mpeg) unchanged. In terms of offered traffic patterns, we used constant interpacket delay, packet bursts, and exponential interpacket delay distributions. In terms of handoffs, we performed two types of handoffs: (a) *hot* switches, in which the mobile host registered with the new `mrouted` before handing off, and (b) *cold* switches, in which the mobile host handed off before registering with the new `mrouted`. In each case, we experimented with handing off between cells, subnets, and switching between interfaces. The introduction of the 100ms delay at *durga* was used in our comparison of MSM-IP with Mobile IP, as described later in this section. The parameters we measured included packet loss upon cold switches and packet duplicates upon hot switches for UDP traffic at the receiving application, and packet traces using `tcpdump` for TCP traffic. In all cases, the measurements were taken for both uplink and downlink.

We report three sets of results in this paper. The first set of results measures packet loss and duplicates for

[1] Note that the change in TCP still assumes that there is only one receiver with the multicast address.

RED HAT 1010
RED HAT V. CAS

UDP traffic with switching between two interfaces connected to the same network. The second set of results measures packet loss and duplicates for UDP traffic with handoff between subnets and switching between two interfaces connected to different subnets (subnet of *radha* to subnet of *atri*). Note that this involves path changes beyond *durga*. The third set of results measures TCP round-trip time variation with switching between two interfaces in different networks.

For the first set of results, we found that for both hot and cold switches within the same network, the effect of both handoff and interface switching is not noticeable even at 1Mbps. At end-to-end throughputs of up to 100Kbps, typically no packets are either lost or duplicated in cold/hot switches. The maximum packet loss observed was 3 packets on the downlink for a cold switch involving bursty traffic at 512 Kbps. Average loss for downlink traffic is 1 packet for a cold switch at 1Mbps, while average loss for the uplink is 1.6 packets for a cold switch at 1 Mbps. For hot switches, we did not observe any duplicates, since the time between advance registration and handoff was negligible. The results of our performance analysis for handoff between cells connected to the same subnet is presented in Table 2.

Mode	Flow	Size	T'put	Loss	Dup
Hot	Down	512	10-512	0	0
Hot	Down	512	1024	1	0
Hot	Up	512	10-1024	0	0
Cold	Up	512	10-100	0	0
Cold	Up	512	512	0.8	0
Cold	Up	512	1024	1.6	0
Cold	Down	512	10-100	0	0
Cold	Down	512	512	3	0
Cold	Down	512	1024	1	0
Hot	Down	1024	10-1024	0	0
Hot	Up	1024	10-1024	0	0
Cold	Up	1024	10-512	0	0
Cold	Up	1024	1024	2	0
Cold	Down	1024	10-100	0	0
Cold	Down	1024	512	0.4	0
Cold	Down	1024	1024	0	0

Table 2: Performance of MSM-IP For Handoff Within The Same Network. The traffic model for all experiments reported in this table was Constant. 'Mode', 'Flow', 'Size' and 'T'put' indicate the switching mode, direction of the flow, packet size (in bytes) and offered UDP throughput (in Kbps) respectively. 'Loss' and 'Dup' indicate the packet loss and duplicates that were measured by the receiving application.

Causing either interface changes or handoffs across two different subnets involves additional overhead since it involves the new multicast router to join the multicast distribution tree (in the case of handoff within the same subnet, the new base station will also receive the multicast packets automatically). We present the measurements when the subnet switching is associated with interface change. The MSM-IP performance is good even for interface switching at relatively good throughputs for wireless networks. For the second set of results, the maximum packet loss we observed was 1, and we recorded a maximum of two duplicates. The performance was studied at throughputs of 100 Kbps and 400 Kbps. The two networks shared a common gateway (*durga*) two hops away. The results of our performance measurements for switching across different networks in presented in Table 3. Note, that the results presented represent not only a handoff across neighboring networks, but also an interface change.

Mode	Flow	T'put	Model	Loss	Dup
Cold	Down	100	Constant	0.25	0
Cold	Up	100	Constant	1	0
Cold	Down	400	Constant	1	0
Cold	Up	400	Constant	1	0
Hot	Down	100	Constant	0	0.75
Hot	Up	100	Constant	1	0
Hot	Down	400	Constant	0	2.5
Hot	Up	400	Constant	1	0
Cold	Down	100	Poisson	0.75	0
Cold	Up	100	Poisson	1	0
Cold	Down	400	Poisson	1	0
Cold	Up	400	Poisson	1	0
Hot	Down	100	Poisson	0	0.75
Hot	Up	100	Poisson	0.75	0
Hot	Down	400	Poisson	0.33	1.75
Hot	Up	400	Poisson	1	0

Table 3: Performance of MSM-IP with Handoff Across Two Different Subnetworks. 'Mode', 'Flow', 'T'put' and 'Model' indicate the switching mode, direction of the flow, offered UDP throughput (in Kbps) and traffic model respectively. 'Loss' and 'Dup' indicate the packet loss and duplicates that were measured by the receiving application.

In order to measure the effect of handoffs and interface change on TCP, we traced the rtt of packets while transmitting at peak rate. An interesting phenomenon that we have observed when performing interface changes and/or handoffs is that for connections with a long delay, TCP over Mobile IP can experience loss of throughput for several seconds. Essentially, packets that were directed to the old interface are lost, and eventually, when the new interface sends duplicate ACKs, fast retransmit and recovery kicks in for a sequence of several lost packets even if the RTO does not expire. The work in [11] addresses this problem and proposes a solution which enables TCP to recognize handoff and interface change, and treat the consequent packet loss differently from congestion, which partially alleviates this problem. In this work, we use TCP unchanged (except for disabling checks for Class D source/destination addresses) over MSM-IP.

In our third set of results, we show the effect of a cold interface switch between two Ethernet interfaces between

RED HAT 1010
RED HAT V. CAS

two subnets. Figure 3 plots the `rtt` value for the sequence of TCP segments. The interface was changed at packet number 350, with the effect that the round trip time went up by about 2ms for about 20 packets before stabilizing to the old value.

When running the WWW browsing, ftp and Mpeg applications, we found no perceptible impact (qualitatively) upon the applications during or immediately after the handoff or interface change. The caveat is that we have only tested MSM-IP in our laboratory environment with handoffs and interface changes between subnets that share gateways two hops away. Further work needs to be done to explore how the performance scales to larger networks.

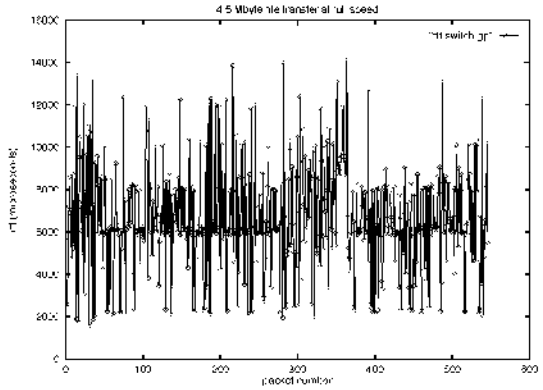


Figure 3: The variation of RTT with switching. The interface change occurred at segment 350, and caused the perceived round trip time of the next 20 segments to go up by approximately 2ms.

5 Key Issues and Problems

The MSM-IP architecture has several advantages and disadvantages as compared to the Mobile IP approach. Among the major advantages are the elegant integration of architectures to support multicasting and mobility, seamless mobility during both handoffs and interface changes, advance registration/reservation, and RSVP-based resource reservation for future Integrated Services networks. Among the major disadvantages are the inability of the state-of-the-art of IP multicasting to provide scalable location management and security/authentication. Over the course of our testbed experimentation, it became clear that the current MSM-IP architecture is still not sufficiently mature enough for widespread deployment. Unfortunately, most of the problems that we encountered in the MSM-IP approach are also also unsolved in the context of IP multicasting currently. While it is true that the basic MSM-IP architecture should be able to support seamless host mobility without any changes in the backbone, several implementation constraints and design issues remain to be solved before an Internet-wide deployment of MSM-IP will be possible. This section identifies the key issues and proposes initial solutions to some of them.

In terms of implementation, the following 6 issues need to be solved: (a) ARP reply processing, (b) TCP support, (c) IGMP membership registration, (d) ICMP message delivery, (e) Multicast TTL initialization in static hosts, and (f) location dependent addressing using DHCP. In terms of the design of the MSM-IP architecture, the following 2 issues need to be addressed: (a) security and authentication, and (b) scalability. We discuss each of these issues below.

1. ARP reply processing: ARP requests from the mobile host will have a multicast source address. The ARP response packet thus has a multicast destination IP address (but unicast link layer interface address). In the standard Linux kernel we used, the ARP response packet is thrown away, causing the host to hang at times. By modifying the kernel to recognize a multicast packet that matches the IP address of the host, we overcame this problem. An alternative solution is to acquire a dynamic location dependent address for use with ARP to over come this problem (item 6).

2. TCP support: Currently, there is no support for TCP over IP-multicasting. In general, supporting a TCP-like protocol over multicasting can be exceedingly complex; however, in our environment, the multicast address corresponding to a mobile host is guaranteed to have only one destination host. In the unmodified TCP version of Linux 2.0.30, we found that when a mobile host initiates a TCP connection request, the SYN packet is acknowledged with a SYN-ACK by correspondent hosts, but the mobile host ignores the SYN-ACK packet since the destination address is a multicast address (`tcp_input()` drops the packet according to RFC 1122 4.2.3.10). The change that is required to the TCP code at the mobile host in order to make it work with multicast mobile host addresses is minimal. However, in several TCP implementations, the correspondent host drops a SYN request if the sender is a multicast address. Supporting TCP connections with such hosts will involve a change to the TCP code in both mobile and correspondent hosts. The emerging interest in reliable multicast applications will push for a variation in TCP that supports multiple communicating hosts, e.g. RMTP [13]. As in the case of other multicasting functionality identified in this paper, this is a superset of the requirements to support mobility within the MSM-IP framework.

3. Initializing the multicast TTL: In the current IP multicasting implementation, the default TTL for a multicast packet is set to 1. Thus, when a correspondent host transmits packets to a mobile host not in its own network, it needs to explicitly set the TTL (otherwise the local multicast router will ignore the packet). Application-transparent support is needed to set the value of the TTL to an appropriate value (as in unicast packets), so that end-user

RED HAT 1010
RED HAT V. CAS

applications can remain unaltered. This problem has more to do with the way DVMRP works than an inherent problem with the MSM-IP approach.

4. IGMP membership registration: Upon handoff or for advance registration, the mobile host needs to register its (multicast) address with the multicast router in its subnet so that the router will begin to receive the multicast packets addressed to the mobile host. Registration is performed through IGMP. There are two problems with the way IGMP works when the mobile host address is a multicast address: (a) at the mobile host, when the mobile host sends the IGMP registration message to the multicast group identified by its address, the address resolution module at the mobile maps the address to the link-level local interface address rather than the link-level multicast address; consequently, the multicast router does not look at the registration message, and (b) even if the above problem is fixed (so that the link-level address is set to the multicast link address), the multicast router is unable to determine on which subnet it must transmit the packets destined for the multicast group. Usually, by looking at the source address of the IGMP request, the router can determine from which subnet the request was made. However, when the mobile host sends a registration request, the source address is a subnet-independent multicast address. As with the ARP problem, acquiring a location-dependent dynamic unicast address solves this problem in a simple way.

5. ICMP message delivery: ICMP messages are issued by intermediate routers when a message or delivery error is detected. These messages are sent back to the original sender of the datagram. In MSM-IP messages from the mobile host have a multicast source address; consequently, intermediate routers will not be able to send ICMP packets back to the mobile host. Of course, tunneled packets by the multicast routers will still have the unicast address of the routers in their headers. Thus, the problem relates to ICMP over the local subnet of the mobile host. We are exploring simple support at the multicast router to overcome this problem as a part of ongoing work.

6. Location Independent Addressing and DHCP: While we propose MSM-IP as the architecture to support effect macro and micro-level host mobility in the Internet, several network management functions require a location dependent address for the mobile host. For these reasons, our support architecture uses DHCP [10] in order to provide dynamic location-specific addresses to mobile hosts. This approach works well for most of the network management related problems we have encountered thus far.

7. Security and Authentication: One of the major issues we need to resolve before the MSM-IP approach reaches maturity is the support for security and authentication in such an environment. Inherently, multicasting is more open to security problems, since any secret or key will need to be shared among all the members of the multicast group. Besides, the shared key may need to be changed each time a host joins or leaves the group. The current design of IP multicasting allows any host to join a multicast group merely by registering with the multicast router in its subnet. While it is well possible to provide secure multicast networks, wherein the multicast routers are isolated from other multicast routers in the Internet, it is much harder to provide for secure multicast groups within the generic framework of IP multicasting.

   Fortunately for us, the MSM-IP approach offers a much simpler context in which to provide security and authentication. In this discussion, our goal is to first provide mechanisms for authentication between communicating entities; data security is an issue that can be handled by encryption or digital signatures at a higher level. In the MSM-IP architecture, the location management hierarchy provides a convenient framework for supporting authentication. Essentially, the location server of a mobile host plays the role of the home agent in Mobile IP from the perspective of authentication. The mobile host and the location server share a secret key which can change over time. In effect, the same authentication framework that is applied to the mobile host, foreign agent and home agent interactions in Mobile IP can be applied to the mobile host, multicast router, and location server interactions in MSM-IP.

8. Scalability: One of the key features of Mobile IP is its scalability. There are no additional constraints in terms of addressing (except for dynamic address allocation where supported) imposed over standard unicast addressing. Home agents and foreign agents need be deployed only in networks that support mobile hosts. Location management is distributed and handled by the individual home networks of mobile hosts. Consequently, Mobile IP substantially retains the laissez-faire nature of the Internet. On the other hand, response to a handoff only occurs when the new care-of address propagates back to either the home agent or the correspondent host (in IPv6). Thus, in terms of support for seamless roaming with ongoing connections, the scale of the connections is likely to be severely restricted. In MSM-IP, management of a shared address pool is a long term issue that needs to be addressed. Multicast routers need to be deployed in every subnet that wishes to support communication with mobile hosts. Potentially, TCP in many flavors of operating systems needs to be changed to accommodate multicast end-point addresses. However, note that none of these problems are unique to mobility support in multicasting architectures. All of the above issues of scale need to be resolved even for standard multicasting services in the Internet. We believe that once this infrastructure is in place, support-

RED HAT 1010
RED HAT V. CAS

ing mobility in such an infrastructure is a trivial issue. The location management hierarchy emulates the approach in Mobile IP and is highly scalable. The multicasting approach for setting up and tearing down routes with receiver initiated joins/prunes and soft state is highly scalable, and can support seamless mobility effectively. Clearly, the benefits of the MSM-IP approach over Mobile IP and its variants are in superior performance, support for seamless mobility, and support for the requirements of future Integrated Services networks.

6 Related Work

In this section we compare the MSM-IP approach to other contemporary approaches to providing mobile host support. A common feature of existing approaches for providing network level support for host mobility is the use of a two tier addressing architecture, consisting of a home address and a care-of address. As described in [3], the key functions of any architecture for mobility support include address translation (mapping of the home address to the care-of address), packet forwarding (the tunneling of packets destined to the home address to the location of the care-of address), and location management (discovery and update of the mobile host's location - typically transparent to the correspondent host). Based on these functionalities, [3] compares the following approaches: Columbia [12], Sony [22], Loose Source Routing [16], Mobile IP [19], and IPv6 [17]. Other approaches to support host mobility include Mosquitonet [5] and Daedalus [20]. We briefly described the Mobile IP approach in Section 2 and compared it with MSM-IP.

The approach in Mosquitonet is very similar to the instance of Mobile IP in which a stripped down version of the foreign agent is co-located with the mobile host. Essentially, a mobile host may acquire a care-of address using a protocol such as DHCP, and then encapsulate or decapsulate packets itself, thereby performing the packet forwarding functionality of the foreign agent.

The approach in Daedalus is to multicast packets from the home agent to the cluster of foreign agents in the neighborhood of the mobile host. As the mobile host moves, this cluster changes. This approach is similar to the MSM-IP approach in its goal of using multicasting to reduce handoff latency. However, it still retains the two-level addressing hierarchy of Mobile IP, and thus incurs the same benefits and overheads (e.g. triangular routing) of Mobile IP. In this approach, multicasting is provided only as a convenience for the sole purpose of reducing packet loss upon handoff. In MSM-IP, multicasting is the *only* form of routing to a mobile host. Our whole point is that once an effective IP multicasting infrastructure is in place, it can be used as-is for supporting mobility. Thus, MSM-IP differs from Daedalus from both the theoretical perspective and the practical issues it needs to address.

The Columbia approach [12] was among the pioneering efforts to support mobility in the Internet. MSM-IP and the Columbia approach share several common goals and features; in fact, MSM-IP can be viewed as the multicasting analogue of the Columbia approach. In the Columbia approach, mobile hosts belong to a virtual mobile network with a distinct network id. A collection of dedicated mobile support routers (MSRs) are used to provide packet forwarding and location management. MSRs communicate with each other by means of tunnels. Mobile host locations are updated and propagated by means of a distributed directory protocol. The key distinction between the Columbia approach and the MSM-IP approach is the use of multicasting in order to reduce handoff packet loss, the ability to advance register and perform resource reservation, and the use of the existing IP multicasting infrastructure to accomplish host mobility.

In the area of advance reservation, a number of recent efforts have proposed the use of multicasting for advance registration and reservation in the neighborhood of the mobile host [14]. The determination of the neighborhood ranges from simply reserving in every neighboring cell, to using sophisticate probabilistic techniques to determine 'shadow clusters', to profiling user movement in order to determine a subset of 'next-predicted-cells'. Again, as in the case of Daedalus, multicasting has been suggested as an enhancement to unicast routing rather than as a replacement of the basic routing scheme itself.

In terms of complementary work in the area of resource reservation, RSVP [24] provides an effective and dynamic resource reservation scheme for multicast communication. Since one of the main goals of RSVP is to provide effective resource reservation mechanisms in the presence of dynamic host joins and prunes, RSVP works well as a resource reservation mechanism in a mobile computing environment. Also, since another main goal of RSVP is to be independent of the underlying multicast routing algorithm, it works well with the MSM-IP architecture. In effect, we can use RSVP without modifications as the resource reservation mechanism for mobile flows. Given that arbitration of scarce and shared wireless networking resources will be a major concern of future Integrated Services mobile computing environments, the seamless interoperation of RSVP with MSM-IP is extremely beneficial.

7 Conclusions

In this paper, we have proposed a new architecture for supporting host mobility in the Internet using IP multicasting as the sole mechanism for routing packets to the mobile hosts. Our approach, called MSM-IP, assigns a unique location independent (multicast) address to a mobile host. Consequently, we eliminate the problem of mapping location independent to location dependent addresses for mobile hosts, which is the key focus of almost every contemporary approach for supporting host mobility in the Internet. Since our mobile hosts are identified by multicast addresses, packets from the correspondent host to the mobile host will be tunneled through a sequence of multicast routers in the multicast distribution tree and reach the mobile host, rather than go through a home network as in Mobile IP.

As a result of using multicasting for supporting host mobility, performing advance registration and having pack-

RED HAT 1010
RED HAT V. CAS

ets delivered to the next cell in advance of a handoff is very natural. Likewise, advance resource reservation and the use of RSVP-like mechanisms for resource reservation are also very natural in our approach. IP multicasting seeks to support efficient joins and prunes of receivers in a multicast group. This helps us perform handoffs in an efficient and graceful manner.

We have implemented a testbed and extensively tested its performance for handoffs and interface changes within and across networks, with hot and cold switches, with TCP, UDP and a variety of communication-intensive applications. Our results indicate that MSM-IP performs very well for all the above tests, and is definitely a viable alternative to contemporary approaches for supporting mobility from the perspective of performance.

However, a number of factors currently prevent the deployment of the IP multicasting infrastructure as-is to support host mobility. Location management is an issue that needs to be resolved effectively - we provide a solution to this problem. Interacting and supporting protocols such as ARP, IGMP, ICMP and TCP (at least for the mobile host) either require changes or modifications to the IP multicasting infrastructure in order to work seamlessly with the routing protocol. We have identified a list of issues that need to be solved, and provided preliminary solutions to most of these issues. However, future work will involve a more detailed study of these issues and attempt to provide better and more scalable solutions to the problems that are currently in the way of allowing an MSM-IP architecture to be widely used for supporting host mobility in the Internet.

References

[1] H. Balakrishnan, V.N. Padmanabhan, S. Seshan, and R.H. Katz. *A Comparison of Mechanisms for Improving TCP Performance over Wireless Links*, **Proceedings of ACM SIGCOMM**, 1996.

[2] A. Ballardie, P. Francis, and J. Crowcroft. *Core Based Trees*, **Proceedings of ACM SIGCOMM**, August 1993.

[3] P. Bhagwat, C. Perkins, and Satish K. Tripathi. *Network Layer Mobility : an Architecture and Survey*, **IEEE Personal Communication** v.3 no. 3, June 1996.

[4] R. Caceres and V. Padmanabhan. *Fast and Scalable Handoffs for Wireless Internetworks*, **Proceedings of ACM Mobicom**, November 1996.

[5] S. Cheshire and M. Baker. *Internet Mobility 4x4*, **Proceedings of ACM SIGCOMM**, 1996.

[6] Y. Dalal and R. Metcalfe. *Reverse Path Forwarding of Broadcast Packets*, **Communications of the ACM**, v. 21, no. 12, December 1978.

[7] S. Deering and J. Hinden. *Internet Protocol, Version 6 (IPv6) Specification*, **RFC 1883**, December 1995.

[8] S. Deering, D. Estrin, **et al**. *Protocol Independent Multicast - Sparse Mode (PIM-SM): Motivation and Architecture*, **draft-ietf-idmr-pim-arch-04.ps**, October 1996.

[9] S. Deering and D. Cheriton. *Multicast Routing in Datagram Internetworks and Extended LANs*, **ACM Transactions on Computer Systems**, May 1990.

[10] R. Droms. *Dynamic Host Configuration Protocol*, **RFC 1541**, October 1993.

[11] M. Stangel, V. Gupta, and V. Bharghavan. *TCP Connection Management for Lossy Mobile Handoff*, In preparation.

[12] J. Ioannidis, D. Duchamp, and G.Q. Maguire Jr.. *IP-based Protocols for Mobile Internetworking*, **Proceedings of ACM SIGCOMM**, September 1991.

[13] J.C. Lin and S. Paul. *RMTP: A Reliable Multicast Transport Protocol*, newblock**Proceedings of IEEE Infocom**, April 1996.

[14] S. Lu and V. Bharghavan. *Adaptive Resource Management Algorithms in Mobile Computing Environments*, **Proceedings of ACM SIGCOMM**, August 1996.

[15] J. Moy. *Multicast Extension to OSPF*, **Internet Draft**, September 1992.

[16] C. Perkins and P. Bhagwat. *A Mobile Networking System based on Internet Protocol*, **IEEE Personal Communications Magezine**, v.1, no. 1, February 1994.

[17] C. Perkins and D. Johnson. *Mobility Support in IPv6*, **draft-ietf-mobileip-ipv6-00.txt**, January 1996.

[18] C. Perkins and D. Johnson. *Mobility Support in IPv6*, **Proceedings of ACM Mobicom**, November 1996.

[19] C. Perkins, Editor. *IP Mobility Support*, **Network Working Group, RFC 2002**, October 1996.

[20] S. Seshan, H. Balakrishnan and R. H. Katz. *Handoffs in Cellular Wireless Networks: The Daedalus Implementation and Experience*, **ACM/Baltzer Journal on Wireless Networks**, 1995.

[21] G. Wright and W. Stevens. *TCP/IP Illustrated, Volume 2*, **Addison Weslay**, 1995.

[22] F. Teraoka, Y. Yokote, and M. Tokoro. *A Network Architecture Providing Host Migration Transparency*, **Proceedings of ACM SIGCOMM**, 1991.

[23] D. Waitzman, S. Deering, and C. Partridge. *Distance Vector Multicast Routing Protocol*, **RFC 1075**, November 1988.

[24] L. Zhang, S. Deering, D. Estrin, S. Shenker, and D. Zappala. *RSVP: A new Resource reSerVation Protocol*, **IEEE Network**, September 1993.

RED HAT 1010
RED HAT V. CAS

TRANSMISSION CONTROL PROTOCOL

DARPA INTERNET PROGRAM

PROTOCOL SPECIFICATION

September 1981

prepared for

Defense Advanced Research Projects Agency
Information Processing Techniques Office
1400 Wilson Boulevard
Arlington, Virginia 22209

by

Information Sciences Institute
University of Southern California
4676 Admiralty Way
Marina del Rey, California 90291

RED HAT 1010
RED HAT V. CAS

TABLE OF CONTENTS

PREFACE iii

1. INTRODUCTION 1

1.1 Motivation 1
1.2 Scope 2
1.3 About This Document 2
1.4 Interfaces 3
1.5 Operation 3

2. PHILOSOPHY 7

2.1 Elements of the Internetwork System 7
2.2 Model of Operation 7
2.3 The Host Environment 8
2.4 Interfaces 9
2.5 Relation to Other Protocols 9
2.6 Reliable Communication 9
2.7 Connection Establishment and Clearing 10
2.8 Data Communication 12
2.9 Precedence and Security 13
2.10 Robustness Principle 13

3. FUNCTIONAL SPECIFICATION 15

3.1 Header Format 15
3.2 Terminology 19
3.3 Sequence Numbers 24
3.4 Establishing a connection 30
3.5 Closing a Connection 37
3.6 Precedence and Security 40
3.7 Data Communication 40
3.8 Interfaces 44
3.9 Event Processing 52

GLOSSARY 79

REFERENCES 85

RED HAT 1010
RED HAT V. CAS

PREFACE

This document describes the DoD Standard Transmission Control Protocol (TCP). There have been nine earlier editions of the ARPA TCP specification on which this standard is based, and the present text draws heavily from them. There have been many contributors to this work both in terms of concepts and in terms of text. This edition clarifies several details and removes the end-of-letter buffer-size adjustments, and redescribes the letter mechanism as a push function.

Jon Postel

Editor

RED HAT 1010
RED HAT V. CAS

RFC: 793
Replaces: RFC 761
IENs: 129, 124, 112, 81,
55, 44, 40, 27, 21, 5

TRANSMISSION CONTROL PROTOCOL

DARPA INTERNET PROGRAM
PROTOCOL SPECIFICATION

1. INTRODUCTION

The Transmission Control Protocol (TCP) is intended for use as a highly reliable host-to-host protocol between hosts in packet-switched computer communication networks, and in interconnected systems of such networks.

This document describes the functions to be performed by the Transmission Control Protocol, the program that implements it, and its interface to programs or users that require its services.

1.1. Motivation

Computer communication systems are playing an increasingly important role in military, government, and civilian environments. This document focuses its attention primarily on military computer communication requirements, especially robustness in the presence of communication unreliability and availability in the presence of congestion, but many of these problems are found in the civilian and government sector as well.

As strategic and tactical computer communication networks are developed and deployed, it is essential to provide means of interconnecting them and to provide standard interprocess communication protocols which can support a broad range of applications. In anticipation of the need for such standards, the Deputy Undersecretary of Defense for Research and Engineering has declared the Transmission Control Protocol (TCP) described herein to be a basis for DoD-wide inter-process communication protocol standardization.

TCP is a connection-oriented, end-to-end reliable protocol designed to fit into a layered hierarchy of protocols which support multi-network applications. The TCP provides for reliable inter-process communication between pairs of processes in host computers attached to distinct but interconnected computer communication networks. Very few assumptions are made as to the reliability of the communication protocols below the TCP layer. TCP assumes it can obtain a simple, potentially unreliable datagram service from the lower level protocols. In principle, the TCP should be able to operate above a wide spectrum of communication systems ranging from hard-wired connections to packet-switched or circuit-switched networks.

RED HAT 1010
RED HAT V. CAS

TCP is based on concepts first described by Cerf and Kahn in [1]. The TCP fits into a layered protocol architecture just above a basic Internet Protocol [2] which provides a way for the TCP to send and receive variable-length segments of information enclosed in internet datagram "envelopes". The internet datagram provides a means for addressing source and destination TCPs in different networks. The internet protocol also deals with any fragmentation or reassembly of the TCP segments required to achieve transport and delivery through multiple networks and interconnecting gateways. The internet protocol also carries information on the precedence, security classification and compartmentation of the TCP segments, so this information can be communicated end-to-end across multiple networks.

Protocol Layering

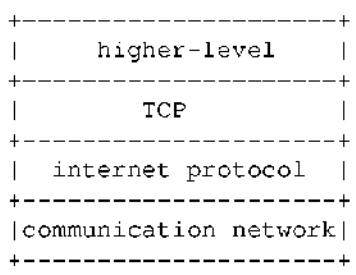


Figure 1

Much of this document is written in the context of TCP implementations which are co-resident with higher level protocols in the host computer. Some computer systems will be connected to networks via front-end computers which house the TCP and internet protocol layers, as well as network specific software. The TCP specification describes an interface to the higher level protocols which appears to be implementable even for the front-end case, as long as a suitable host-to-front end protocol is implemented.

1.2. Scope

The TCP is intended to provide a reliable process-to-process communication service in a multinetwork environment. The TCP is intended to be a host-to-host protocol in common use in multiple networks.

1.3. About this Document

This document represents a specification of the behavior required of any TCP implementation, both in its interactions with higher level protocols and in its interactions with other TCPs. The rest of this

RED HAT 1010
RED HAT V. CAS

section offers a very brief view of the protocol interfaces and operation. Section 2 summarizes the philosophical basis for the TCP design. Section 3 offers both a detailed description of the actions required of TCP when various events occur (arrival of new segments, user calls, errors, etc.) and the details of the formats of TCP segments.

1.4. Interfaces

The TCP interfaces on one side to user or application processes and on the other side to a lower level protocol such as Internet Protocol.

The interface between an application process and the TCP is illustrated in reasonable detail. This interface consists of a set of calls much like the calls an operating system provides to an application process for manipulating files. For example, there are calls to open and close connections and to send and receive data on established connections. It is also expected that the TCP can asynchronously communicate with application programs. Although considerable freedom is permitted to TCP implementors to design interfaces which are appropriate to a particular operating system environment, a minimum functionality is required at the TCP/user interface for any valid implementation.

The interface between TCP and lower level protocol is essentially unspecified except that it is assumed there is a mechanism whereby the two levels can asynchronously pass information to each other. Typically, one expects the lower level protocol to specify this interface. TCP is designed to work in a very general environment of interconnected networks. The lower level protocol which is assumed throughout this document is the Internet Protocol [2].

1.5. Operation

As noted above, the primary purpose of the TCP is to provide reliable, securable logical circuit or connection service between pairs of processes. To provide this service on top of a less reliable internet communication system requires facilities in the following areas:

- Basic Data Transfer
- Reliability
- Flow Control
- Multiplexing
- Connections
- Precedence and Security

The basic operation of the TCP in each of these areas is described in the following paragraphs.

RED HAT 1010
RED HAT V. CAS

Basic Data Transfer:

The TCP is able to transfer a continuous stream of octets in each direction between its users by packaging some number of octets into segments for transmission through the internet system. In general, the TCPs decide when to block and forward data at their own convenience.

Sometimes users need to be sure that all the data they have submitted to the TCP has been transmitted. For this purpose a push function is defined. To assure that data submitted to a TCP is actually transmitted the sending user indicates that it should be pushed through to the receiving user. A push causes the TCPs to promptly forward and deliver data up to that point to the receiver. The exact push point might not be visible to the receiving user and the push function does not supply a record boundary marker.

Reliability:

The TCP must recover from data that is damaged, lost, duplicated, or delivered out of order by the internet communication system. This is achieved by assigning a sequence number to each octet transmitted, and requiring a positive acknowledgment (ACK) from the receiving TCP. If the ACK is not received within a timeout interval, the data is retransmitted. At the receiver, the sequence numbers are used to correctly order segments that may be received out of order and to eliminate duplicates. Damage is handled by adding a checksum to each segment transmitted, checking it at the receiver, and discarding damaged segments.

As long as the TCPs continue to function properly and the internet system does not become completely partitioned, no transmission errors will affect the correct delivery of data. TCP recovers from internet communication system errors.

Flow Control:

TCP provides a means for the receiver to govern the amount of data sent by the sender. This is achieved by returning a "window" with every ACK indicating a range of acceptable sequence numbers beyond the last segment successfully received. The window indicates an allowed number of octets that the sender may transmit before receiving further permission.

RED HAT 1010
RED HAT V. CAS

Multiplexing:

To allow for many processes within a single Host to use TCP communication facilities simultaneously, the TCP provides a set of addresses or ports within each host. Concatenated with the network and host addresses from the internet communication layer, this forms a socket. A pair of sockets uniquely identifies each connection. That is, a socket may be simultaneously used in multiple connections.

The binding of ports to processes is handled independently by each Host. However, it proves useful to attach frequently used processes (e.g., a "logger" or timesharing service) to fixed sockets which are made known to the public. These services can then be accessed through the known addresses. Establishing and learning the port addresses of other processes may involve more dynamic mechanisms.

Connections:

The reliability and flow control mechanisms described above require that TCPs initialize and maintain certain status information for each data stream. The combination of this information, including sockets, sequence numbers, and window sizes, is called a connection. Each connection is uniquely specified by a pair of sockets identifying its two sides.

When two processes wish to communicate, their TCP's must first establish a connection (initialize the status information on each side). When their communication is complete, the connection is terminated or closed to free the resources for other uses.

Since connections must be established between unreliable hosts and over the unreliable internet communication system, a handshake mechanism with clock-based sequence numbers is used to avoid erroneous initialization of connections.

Precedence and Security:

The users of TCP may indicate the security and precedence of their communication. Provision is made for default values to be used when these features are not needed.

RED HAT 1010
RED HAT V. CAS

2. PHILOSOPHY

2.1. Elements of the Internetwork System

The internetwork environment consists of hosts connected to networks which are in turn interconnected via gateways. It is assumed here that the networks may be either local networks (e.g., the ETHERNET) or large networks (e.g., the ARPANET), but in any case are based on packet switching technology. The active agents that produce and consume messages are processes. Various levels of protocols in the networks, the gateways, and the hosts support an interprocess communication system that provides two-way data flow on logical connections between process ports.

The term packet is used generically here to mean the data of one transaction between a host and its network. The format of data blocks exchanged within the a network will generally not be of concern to us.

Hosts are computers attached to a network, and from the communication network's point of view, are the sources and destinations of packets. Processes are viewed as the active elements in host computers (in accordance with the fairly common definition of a process as a program in execution). Even terminals and files or other I/O devices are viewed as communicating with each other through the use of processes. Thus, all communication is viewed as inter-process communication.

Since a process may need to distinguish among several communication streams between itself and another process (or processes), we imagine that each process may have a number of ports through which it communicates with the ports of other processes.

2.2. Model of Operation

Processes transmit data by calling on the TCP and passing buffers of data as arguments. The TCP packages the data from these buffers into segments and calls on the internet module to transmit each segment to the destination TCP. The receiving TCP places the data from a segment into the receiving user's buffer and notifies the receiving user. The TCPs include control information in the segments which they use to ensure reliable ordered data transmission.

The model of internet communication is that there is an internet protocol module associated with each TCP which provides an interface to the local network. This internet module packages TCP segments inside internet datagrams and routes these datagrams to a destination internet module or intermediate gateway. To transmit the datagram through the local network, it is embedded in a local network packet.

The packet switches may perform further packaging, fragmentation, or

RED HAT 1010
RED HAT V. CAS

other operations to achieve the delivery of the local packet to the destination internet module.

At a gateway between networks, the internet datagram is "unwrapped" from its local packet and examined to determine through which network the internet datagram should travel next. The internet datagram is then "wrapped" in a local packet suitable to the next network and routed to the next gateway, or to the final destination.

A gateway is permitted to break up an internet datagram into smaller internet datagram fragments if this is necessary for transmission through the next network. To do this, the gateway produces a set of internet datagrams; each carrying a fragment. Fragments may be further broken into smaller fragments at subsequent gateways. The internet datagram fragment format is designed so that the destination internet module can reassemble fragments into internet datagrams.

A destination internet module unwraps the segment from the datagram (after reassembling the datagram, if necessary) and passes it to the destination TCP.

This simple model of the operation glosses over many details. One important feature is the type of service. This provides information to the gateway (or internet module) to guide it in selecting the service parameters to be used in traversing the next network. Included in the type of service information is the precedence of the datagram. Datagrams may also carry security information to permit host and gateways that operate in multilevel secure environments to properly segregate datagrams for security considerations.

2.3. The Host Environment

The TCP is assumed to be a module in an operating system. The users access the TCP much like they would access the file system. The TCP may call on other operating system functions, for example, to manage data structures. The actual interface to the network is assumed to be controlled by a device driver module. The TCP does not call on the network device driver directly, but rather calls on the internet datagram protocol module which may in turn call on the device driver.

The mechanisms of TCP do not preclude implementation of the TCP in a front-end processor. However, in such an implementation, a host-to-front-end protocol must provide the functionality to support the type of TCP-user interface described in this document.

RED HAT 1010
RED HAT V. CAS

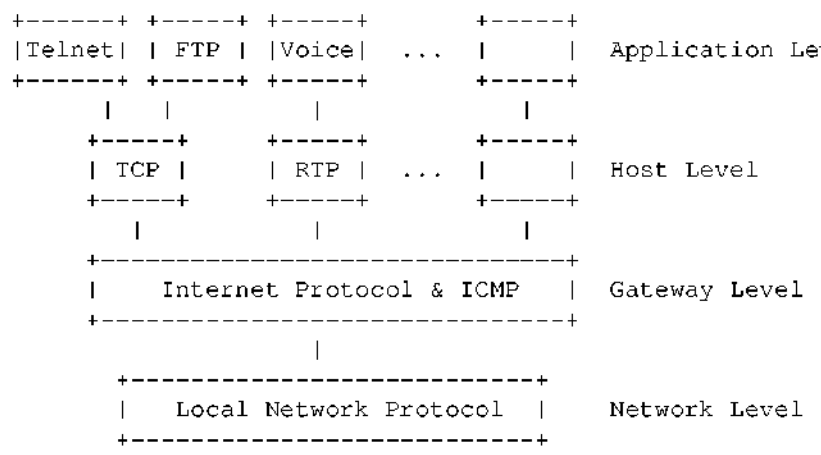
2.4. Interfaces

The TCP/user interface provides for calls made by the user on the TCP to OPEN or CLOSE a connection, to SEND or RECEIVE data, or to obtain STATUS about a connection. These calls are like other calls from user programs on the operating system, for example, the calls to open, read from, and close a file.

The TCP/internet interface provides calls to send and receive datagrams addressed to TCP modules in hosts anywhere in the internet system. These calls have parameters for passing the address, type of service, precedence, security, and other control information.

2.5. Relation to Other Protocols

The following diagram illustrates the place of the TCP in the protocol hierarchy:



Protocol Relationships

Figure 2.

It is expected that the TCP will be able to support higher level protocols efficiently. It should be easy to interface higher level protocols like the ARPANET Telnet or AUTODIN II THP to the TCP.

2.6. Reliable Communication

A stream of data sent on a TCP connection is delivered reliably and in order at the destination.

RED HAT 1010
RED HAT V. CAS

Transmission is made reliable via the use of sequence numbers and acknowledgments. Conceptually, each octet of data is assigned a sequence number. The sequence number of the first octet of data in a segment is transmitted with that segment and is called the segment sequence number. Segments also carry an acknowledgment number which is the sequence number of the next expected data octet of transmissions in the reverse direction. When the TCP transmits a segment containing data, it puts a copy on a retransmission queue and starts a timer; when the acknowledgment for that data is received, the segment is deleted from the queue. If the acknowledgment is not received before the timer runs out, the segment is retransmitted.

An acknowledgment by TCP does not guarantee that the data has been delivered to the end user, but only that the receiving TCP has taken the responsibility to do so.

To govern the flow of data between TCPs, a flow control mechanism is employed. The receiving TCP reports a "window" to the sending TCP. This window specifies the number of octets, starting with the acknowledgment number, that the receiving TCP is currently prepared to receive.

2.7. Connection Establishment and Clearing

To identify the separate data streams that a TCP may handle, the TCP provides a port identifier. Since port identifiers are selected independently by each TCP they might not be unique. To provide for unique addresses within each TCP, we concatenate an internet address identifying the TCP with a port identifier to create a socket which will be unique throughout all networks connected together.

A connection is fully specified by the pair of sockets at the ends. A local socket may participate in many connections to different foreign sockets. A connection can be used to carry data in both directions, that is, it is "full duplex".

TCPs are free to associate ports with processes however they choose. However, several basic concepts are necessary in any implementation. There must be well-known sockets which the TCP associates only with the "appropriate" processes by some means. We envision that processes may "own" ports, and that processes can initiate connections only on the ports they own. (Means for implementing ownership is a local issue, but we envision a Request Port user command, or a method of uniquely allocating a group of ports to a given process, e.g., by associating the high order bits of a port name with a given process.)

A connection is specified in the OPEN call by the local port and foreign socket arguments. In return, the TCP supplies a (short) local

RED HAT 1010
RED HAT V. CAS

connection name by which the user refers to the connection in subsequent calls. There are several things that must be remembered about a connection. To store this information we imagine that there is a data structure called a Transmission Control Block (TCB). One implementation strategy would have the local connection name be a pointer to the TCB for this connection. The OPEN call also specifies whether the connection establishment is to be actively pursued, or to be passively waited for.

A passive OPEN request means that the process wants to accept incoming connection requests rather than attempting to initiate a connection. Often the process requesting a passive OPEN will accept a connection request from any caller. In this case a foreign socket of all zeros is used to denote an unspecified socket. Unspecified foreign sockets are allowed only on passive OPENs.

A service process that wished to provide services for unknown other processes would issue a passive OPEN request with an unspecified foreign socket. Then a connection could be made with any process that requested a connection to this local socket. It would help if this local socket were known to be associated with this service.

Well-known sockets are a convenient mechanism for a priori associating a socket address with a standard service. For instance, the "Telnet-Server" process is permanently assigned to a particular socket, and other sockets are reserved for File Transfer, Remote Job Entry, Text Generator, Echoer, and Sink processes (the last three being for test purposes). A socket address might be reserved for access to a "Look-Up" service which would return the specific socket at which a newly created service would be provided. The concept of a well-known socket is part of the TCP specification, but the assignment of sockets to services is outside this specification. (See [4].)

Processes can issue passive OPENs and wait for matching active OPENs from other processes and be informed by the TCP when connections have been established. Two processes which issue active OPENs to each other at the same time will be correctly connected. This flexibility is critical for the support of distributed computing in which components act asynchronously with respect to each other.

There are two principal cases for matching the sockets in the local passive OPENs and an foreign active OPENs. In the first case, the local passive OPENs has fully specified the foreign socket. In this case, the match must be exact. In the second case, the local passive OPENs has left the foreign socket unspecified. In this case, any foreign socket is acceptable as long as the local sockets match. Other possibilities include partially restricted matches.

RED HAT 1010
RED HAT V. CAS

If there are several pending passive OPENs (recorded in TCBs) with the same local socket, an foreign active OPEN will be matched to a TCB with the specific foreign socket in the foreign active OPEN, if such a TCB exists, before selecting a TCB with an unspecified foreign socket.

The procedures to establish connections utilize the synchronize (SYN) control flag and involves an exchange of three messages. This exchange has been termed a three-way hand shake [3].

A connection is initiated by the rendezvous of an arriving segment containing a SYN and a waiting TCB entry each created by a user OPEN command. The matching of local and foreign sockets determines when a connection has been initiated. The connection becomes "established" when sequence numbers have been synchronized in both directions.

The clearing of a connection also involves the exchange of segments, in this case carrying the FIN control flag.

2.8. Data Communication

The data that flows on a connection may be thought of as a stream of octets. The sending user indicates in each SEND call whether the data in that call (and any preceeding calls) should be immediately pushed through to the receiving user by the setting of the PUSH flag.

A sending TCP is allowed to collect data from the sending user and to send that data in segments at its own convenience, until the push function is signaled, then it must send all unsent data. When a receiving TCP sees the PUSH flag, it must not wait for more data from the sending TCP before passing the data to the receiving process.

There is no necessary relationship between push functions and segment boundaries. The data in any particular segment may be the result of a single SEND call, in whole or part, or of multiple SEND calls.

The purpose of push function and the PUSH flag is to push data through from the sending user to the receiving user. It does not provide a record service.

There is a coupling between the push function and the use of buffers of data that cross the TCP/user interface. Each time a PUSH flag is associated with data placed into the receiving user's buffer, the buffer is returned to the user for processing even if the buffer is not filled. If data arrives that fills the user's buffer before a PUSH is seen, the data is passed to the user in buffer size units.

TCP also provides a means to communicate to the receiver of data that at some point further along in the data stream than the receiver is

RED HAT 1010
RED HAT V. CAS

currently reading there is urgent data. TCP does not attempt to define what the user specifically does upon being notified of pending urgent data, but the general notion is that the receiving process will take action to process the urgent data quickly.

2.9. Precedence and Security

The TCP makes use of the internet protocol type of service field and security option to provide precedence and security on a per connection basis to TCP users. Not all TCP modules will necessarily function in a multilevel secure environment; some may be limited to unclassified use only, and others may operate at only one security level and compartment. Consequently, some TCP implementations and services to users may be limited to a subset of the multilevel secure case.

TCP modules which operate in a multilevel secure environment must properly mark outgoing segments with the security, compartment, and precedence. Such TCP modules must also provide to their users or higher level protocols such as Telnet or THP an interface to allow them to specify the desired security level, compartment, and precedence of connections.

2.10. Robustness Principle

TCP implementations will follow a general principle of robustness: be conservative in what you do, be liberal in what you accept from others.

RED HAT 1010
RED HAT V. CAS

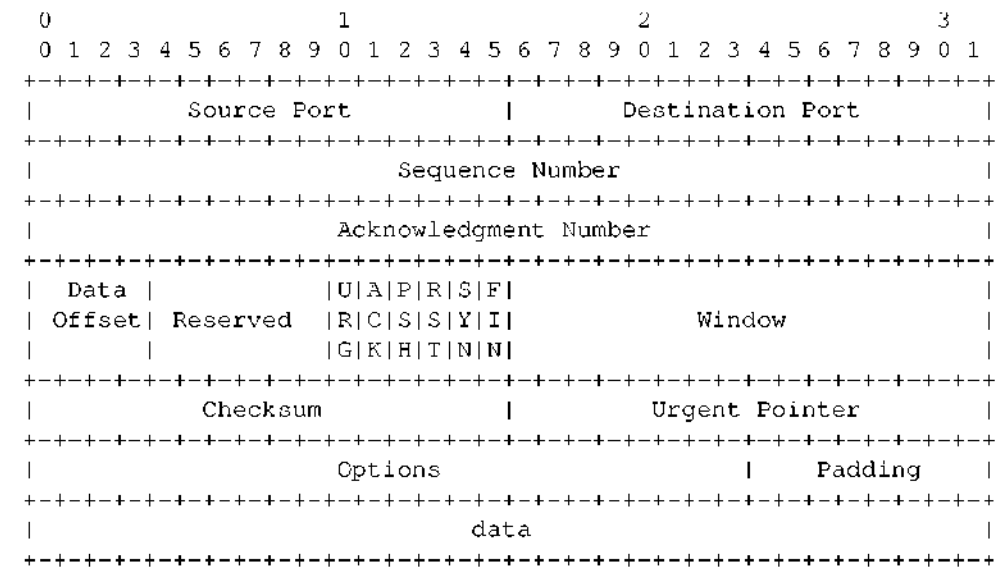
RED HAT 1010
RED HAT V. CAS

3. FUNCTIONAL SPECIFICATION

3.1. Header Format

TCP segments are sent as internet datagrams. The Internet Protocol header carries several information fields, including the source and destination host addresses [2]. A TCP header follows the internet header, supplying information specific to the TCP protocol. This division allows for the existence of host level protocols other than TCP.

TCP Header Format



TCP Header Format

Note that one tick mark represents one bit position.

Figure 3.

Source Port: 16 bits

The source port number.

Destination Port: 16 bits

The destination port number.

RED HAT 1010
RED HAT V. CAS

Sequence Number: 32 bits

The sequence number of the first data octet in this segment (except when SYN is present). If SYN is present the sequence number is the initial sequence number (ISN) and the first data octet is ISN+1.

Acknowledgment Number: 32 bits

If the ACK control bit is set this field contains the value of the next sequence number the sender of the segment is expecting to receive. Once a connection is established this is always sent.

Data Offset: 4 bits

The number of 32 bit words in the TCP Header. This indicates where the data begins. The TCP header (even one including options) is an integral number of 32 bits long.

Reserved: 6 bits

Reserved for future use. Must be zero.

Control Bits: 6 bits (from left to right):

URG: Urgent Pointer field significant
ACK: Acknowledgment field significant
PSH: Push Function
RST: Reset the connection
SYN: Synchronize sequence numbers
FIN: No more data from sender

Window: 16 bits

The number of data octets beginning with the one indicated in the acknowledgment field which the sender of this segment is willing to accept.

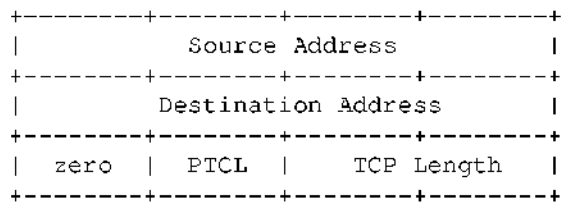
Checksum: 16 bits

The checksum field is the 16 bit one's complement of the one's complement sum of all 16 bit words in the header and text. If a segment contains an odd number of header and text octets to be checksummed, the last octet is padded on the right with zeros to form a 16 bit word for checksum purposes. The pad is not transmitted as part of the segment. While computing the checksum, the checksum field itself is replaced with zeros.

The checksum also covers a 96 bit pseudo header conceptually

RED HAT 1010
RED HAT V. CAS

prefixed to the TCP header. This pseudo header contains the Source Address, the Destination Address, the Protocol, and TCP length. This gives the TCP protection against misrouted segments. This information is carried in the Internet Protocol and is transferred across the TCP/Network interface in the arguments or results of calls by the TCP on the IP.



The TCP Length is the TCP header length plus the data length in octets (this is not an explicitly transmitted quantity, but is computed), and it does not count the 12 octets of the pseudo header.

Urgent Pointer: 16 bits

This field communicates the current value of the urgent pointer as a positive offset from the sequence number in this segment. The urgent pointer points to the sequence number of the octet following the urgent data. This field is only be interpreted in segments with the URG control bit set.

Options: variable

Options may occupy space at the end of the TCP header and are a multiple of 8 bits in length. All options are included in the checksum. An option may begin on any octet boundary. There are two cases for the format of an option:

Case 1: A single octet of option-kind.

Case 2: An octet of option-kind, an octet of option-length, and the actual option-data octets.

The option-length counts the two octets of option-kind and option-length as well as the option-data octets.

Note that the list of options may be shorter than the data offset field might imply. The content of the header beyond the End-of-Option option must be header padding (i.e., zero).

A TCP must implement all options.

RED HAT 1010
RED HAT V. CAS

Currently defined options include (kind indicated in octal):

Kind	Length	Meaning
0	-	End of option list.
1	-	No-Operation.
2	4	Maximum Segment Size.

Specific Option Definitions

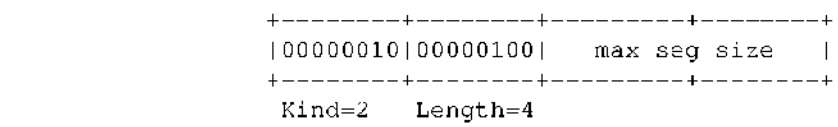
End of Option List

This option code indicates the end of the option list. This might not coincide with the end of the TCP header according to the Data Offset field. This is used at the end of all options, not the end of each option, and need only be used if the end of the options would not otherwise coincide with the end of the TCP header.

No-Operation

This option code may be used between options, for example, to align the beginning of a subsequent option on a word boundary. There is no guarantee that senders will use this option, so receivers must be prepared to process options even if they do not begin on a word boundary.

Maximum Segment Size



RED HAT 1010
RED HAT V. CAS

Maximum Segment Size Option Data: 16 bits

If this option is present, then it communicates the maximum receive segment size at the TCP which sends this segment. This field must only be sent in the initial connection request (i.e., in segments with the SYN control bit set). If this option is not used, any segment size is allowed.

Padding: variable

The TCP header padding is used to ensure that the TCP header ends and data begins on a 32 bit boundary. The padding is composed of zeros.

3.2. Terminology

Before we can discuss very much about the operation of the TCP we need to introduce some detailed terminology. The maintenance of a TCP connection requires the remembering of several variables. We conceive of these variables being stored in a connection record called a Transmission Control Block or TCB. Among the variables stored in the TCB are the local and remote socket numbers, the security and precedence of the connection, pointers to the user's send and receive buffers, pointers to the retransmit queue and to the current segment. In addition several variables relating to the send and receive sequence numbers are stored in the TCB.

Send Sequence Variables

```
SND.UNA - send unacknowledged
SND.NXT - send next
SND.WND - send window
SND.UP  - send urgent pointer
SND.WL1 - segment sequence number used for last window update
SND.WL2 - segment acknowledgment number used for last window
          update
ISS     - initial send sequence number
```

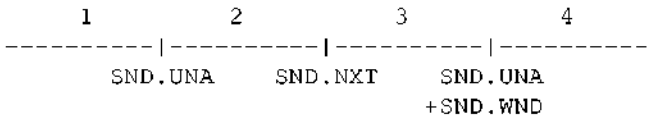
Receive Sequence Variables

```
RCV.NXT - receive next
RCV.WND - receive window
RCV.UP  - receive urgent pointer
IRS     - initial receive sequence number
```


RED HAT 1010
RED HAT V. CAS

The following diagrams may help to relate some of these variables to the sequence space.

Send Sequence Space



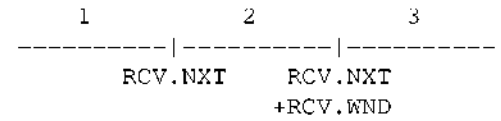
1 - old sequence numbers which have been acknowledged
2 - sequence numbers of unacknowledged data
3 - sequence numbers allowed for new data transmission
4 - future sequence numbers which are not yet allowed

Send Sequence Space

Figure 4.

The send window is the portion of the sequence space labeled 3 in figure 4.

Receive Sequence Space



1 - old sequence numbers which have been acknowledged
2 - sequence numbers allowed for new reception
3 - future sequence numbers which are not yet allowed

Receive Sequence Space

Figure 5.

The receive window is the portion of the sequence space labeled 2 in figure 5.

There are also some variables used frequently in the discussion that take their values from the fields of the current segment.

RED HAT 1010
RED HAT V. CAS

Current Segment Variables

SEG.SEQ - segment sequence number
SEG.ACK - segment acknowledgment number
SEG.LEN - segment length
SEG.WND - segment window
SEG.UP - segment urgent pointer
SEG.PRC - segment precedence value

A connection progresses through a series of states during its lifetime. The states are: LISTEN, SYN-SENT, SYN-RECEIVED, ESTABLISHED, FIN-WAIT-1, FIN-WAIT-2, CLOSE-WAIT, CLOSING, LAST-ACK, TIME-WAIT, and the fictional state CLOSED. CLOSED is fictional because it represents the state when there is no TCB, and therefore, no connection. Briefly the meanings of the states are:

LISTEN - represents waiting for a connection request from any remote TCP and port.

SYN-SENT - represents waiting for a matching connection request after having sent a connection request.

SYN-RECEIVED - represents waiting for a confirming connection request acknowledgment after having both received and sent a connection request.

ESTABLISHED - represents an open connection, data received can be delivered to the user. The normal state for the data transfer phase of the connection.

FIN-WAIT-1 - represents waiting for a connection termination request from the remote TCP, or an acknowledgment of the connection termination request previously sent.

FIN-WAIT-2 - represents waiting for a connection termination request from the remote TCP.

CLOSE-WAIT - represents waiting for a connection termination request from the local user.

CLOSING - represents waiting for a connection termination request acknowledgment from the remote TCP.

LAST-ACK - represents waiting for an acknowledgment of the connection termination request previously sent to the remote TCP (which includes an acknowledgment of its connection termination request).

RED HAT 1010
RED HAT V. CAS

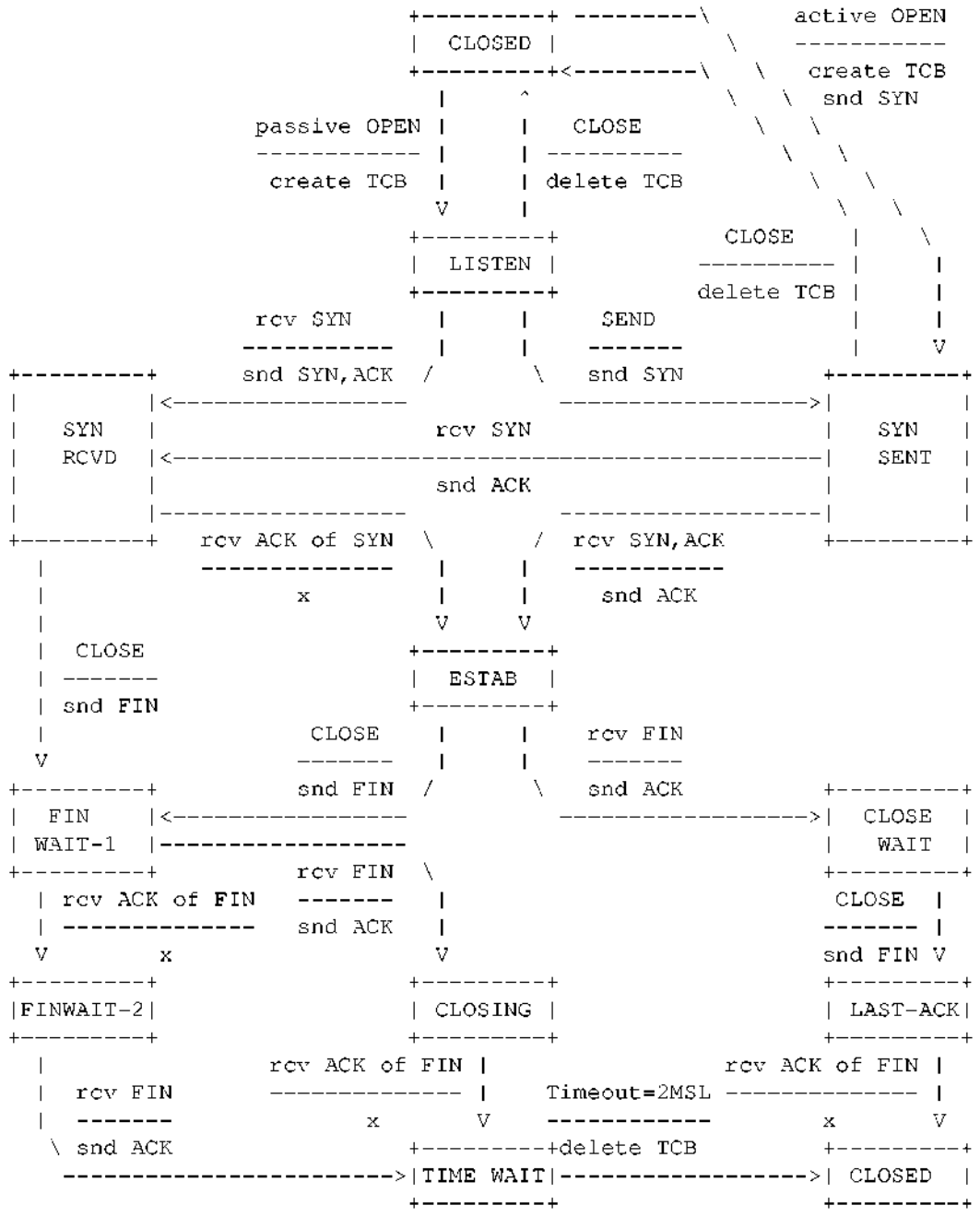
TIME-WAIT - represents waiting for enough time to pass to be sure the remote TCP received the acknowledgment of its connection termination request.

CLOSED - represents no connection state at all.

A TCP connection progresses from one state to another in response to events. The events are the user calls, OPEN, SEND, RECEIVE, CLOSE, ABORT, and STATUS; the incoming segments, particularly those containing the SYN, ACK, RST and FIN flags; and timeouts.

The state diagram in figure 6 illustrates only state changes, together with the causing events and resulting actions, but addresses neither error conditions nor actions which are not connected with state changes. In a later section, more detail is offered with respect to the reaction of the TCP to events.

NOTE BENE: this diagram is only a summary and must not be taken as the total specification.



TCP Connection State Diagram
Figure 6.

3.3. Sequence Numbers

A fundamental notion in the design is that every octet of data sent over a TCP connection has a sequence number. Since every octet is sequenced, each of them can be acknowledged. The acknowledgment mechanism employed is cumulative so that an acknowledgment of sequence number X indicates that all octets up to but not including X have been received. This mechanism allows for straight-forward duplicate detection in the presence of retransmission. Numbering of octets within a segment is that the first data octet immediately following the header is the lowest numbered, and the following octets are numbered consecutively.

It is essential to remember that the actual sequence number space is finite, though very large. This space ranges from 0 to 2**32 - 1. Since the space is finite, all arithmetic dealing with sequence numbers must be performed modulo 2**32. This unsigned arithmetic preserves the relationship of sequence numbers as they cycle from 2**32 - 1 to 0 again. There are some subtleties to computer modulo arithmetic, so great care should be taken in programming the comparison of such values. The symbol "=<" means "less than or equal" (modulo 2**32).

The typical kinds of sequence number comparisons which the TCP must perform include:

(a) Determining that an acknowledgment refers to some sequence number sent but not yet acknowledged.

(b) Determining that all sequence numbers occupied by a segment have been acknowledged (e.g., to remove the segment from a retransmission queue).

(c) Determining that an incoming segment contains sequence numbers which are expected (i.e., that the segment "overlaps" the receive window).

RED HAT 1010
RED HAT V. CAS

In response to sending data the TCP will receive acknowledgments. The following comparisons are needed to process the acknowledgments.

SND.UNA = oldest unacknowledged sequence number

SND.NXT = next sequence number to be sent

SEG.ACK = acknowledgment from the receiving TCP (next sequence number expected by the receiving TCP)

SEG.SEQ = first sequence number of a segment

SEG.LEN = the number of octets occupied by the data in the segment (counting SYN and FIN)

SEG.SEQ+SEG.LEN-1 = last sequence number of a segment

A new acknowledgment (called an "acceptable ack"), is one for which the inequality below holds:

SND.UNA < SEG.ACK =< SND.NXT

A segment on the retransmission queue is fully acknowledged if the sum of its sequence number and length is less or equal than the acknowledgment value in the incoming segment.

When data is received the following comparisons are needed:

RCV.NXT = next sequence number expected on an incoming segments, and is the left or lower edge of the receive window

RCV.NXT+RCV.WND-1 = last sequence number expected on an incoming segment, and is the right or upper edge of the receive window

SEG.SEQ = first sequence number occupied by the incoming segment

SEG.SEQ+SEG.LEN-1 = last sequence number occupied by the incoming segment

A segment is judged to occupy a portion of valid receive sequence space if

RCV.NXT =< SEG.SEQ < RCV.NXT+RCV.WND

or

RCV.NXT =< SEG.SEQ+SEG.LEN-1 < RCV.NXT+RCV.WND

RED HAT 1010
RED HAT V. CAS

The first part of this test checks to see if the beginning of the segment falls in the window, the second part of the test checks to see if the end of the segment falls in the window; if the segment passes either part of the test it contains data in the window.

Actually, it is a little more complicated than this. Due to zero windows and zero length segments, we have four cases for the acceptability of an incoming segment:

Segment Length	Receive Window	Test
0	0	SEG.SEQ = RCV.NXT
0	>0	RCV.NXT =< SEG.SEQ < RCV.NXT+RCV.WND
>0	0	not acceptable
>0	>0	RCV.NXT =< SEG.SEQ < RCV.NXT+RCV.WND or RCV.NXT =< SEG.SEQ+SEG.LEN-1 < RCV.NXT+RCV.WND

Note that when the receive window is zero no segments should be acceptable except ACK segments. Thus, it is be possible for a TCP to maintain a zero receive window while transmitting data and receiving ACKs. However, even when the receive window is zero, a TCP must process the RST and URG fields of all incoming segments.

We have taken advantage of the numbering scheme to protect certain control information as well. This is achieved by implicitly including some control flags in the sequence space so they can be retransmitted and acknowledged without confusion (i.e., one and only one copy of the control will be acted upon). Control information is not physically carried in the segment data space. Consequently, we must adopt rules for implicitly assigning sequence numbers to control. The SYN and FIN are the only controls requiring this protection, and these controls are used only at connection opening and closing. For sequence number purposes, the SYN is considered to occur before the first actual data octet of the segment in which it occurs, while the FIN is considered to occur after the last actual data octet in a segment in which it occurs. The segment length (SEG.LEN) includes both data and sequence space occupying controls. When a SYN is present then SEG.SEQ is the sequence number of the SYN.

Initial Sequence Number Selection

The protocol places no restriction on a particular connection being used over and over again. A connection is defined by a pair of sockets. New instances of a connection will be referred to as incarnations of the connection. The problem that arises from this is -- "how does the TCP identify duplicate segments from previous incarnations of the connection?" This problem becomes apparent if the connection is being opened and closed in quick succession, or if the connection breaks with loss of memory and is then reestablished.

To avoid confusion we must prevent segments from one incarnation of a connection from being used while the same sequence numbers may still be present in the network from an earlier incarnation. We want to assure this, even if a TCP crashes and loses all knowledge of the sequence numbers it has been using. When new connections are created, an initial sequence number (ISN) generator is employed which selects a new 32 bit ISN. The generator is bound to a (possibly fictitious) 32 bit clock whose low order bit is incremented roughly every 4 microseconds. Thus, the ISN cycles approximately every 4.55 hours. Since we assume that segments will stay in the network no more than the Maximum Segment Lifetime (MSL) and that the MSL is less than 4.55 hours we can reasonably assume that ISN's will be unique.

For each connection there is a send sequence number and a receive sequence number. The initial send sequence number (ISS) is chosen by the data sending TCP, and the initial receive sequence number (IRS) is learned during the connection establishing procedure.

For a connection to be established or initialized, the two TCPs must synchronize on each other's initial sequence numbers. This is done in an exchange of connection establishing segments carrying a control bit called "SYN" (for synchronize) and the initial sequence numbers. As a shorthand, segments carrying the SYN bit are also called "SYNs". Hence, the solution requires a suitable mechanism for picking an initial sequence number and a slightly involved handshake to exchange the ISN's.

The synchronization requires each side to send it's own initial sequence number and to receive a confirmation of it in acknowledgment from the other side. Each side must also receive the other side's initial sequence number and send a confirming acknowledgment.

1) A --> B SYN my sequence number is X
2) A <-- B ACK your sequence number is X
3) A <-- B SYN my sequence number is Y
4) A --> B ACK your sequence number is Y

RED HAT 1010
RED HAT V. CAS

Because steps 2 and 3 can be combined in a single message this is called the three way (or three message) handshake.

A three way handshake is necessary because sequence numbers are not tied to a global clock in the network, and TCPs may have different mechanisms for picking the ISN's. The receiver of the first SYN has no way of knowing whether the segment was an old delayed one or not, unless it remembers the last sequence number used on the connection (which is not always possible), and so it must ask the sender to verify this SYN. The three way handshake and the advantages of a clock-driven scheme are discussed in [3].

Knowing When to Keep Quiet

To be sure that a TCP does not create a segment that carries a sequence number which may be duplicated by an old segment remaining in the network, the TCP must keep quiet for a maximum segment lifetime (MSL) before assigning any sequence numbers upon starting up or recovering from a crash in which memory of sequence numbers in use was lost. For this specification the MSL is taken to be 2 minutes. This is an engineering choice, and may be changed if experience indicates it is desirable to do so. Note that if a TCP is reinitialized in some sense, yet retains its memory of sequence numbers in use, then it need not wait at all; it must only be sure to use sequence numbers larger than those recently used.

The TCP Quiet Time Concept

This specification provides that hosts which "crash" without retaining any knowledge of the last sequence numbers transmitted on each active (i.e., not closed) connection shall delay emitting any TCP segments for at least the agreed Maximum Segment Lifetime (MSL) in the internet system of which the host is a part. In the paragraphs below, an explanation for this specification is given. TCP implementors may violate the "quiet time" restriction, but only at the risk of causing some old data to be accepted as new or new data rejected as old duplicated by some receivers in the internet system.

TCPs consume sequence number space each time a segment is formed and entered into the network output queue at a source host. The duplicate detection and sequencing algorithm in the TCP protocol relies on the unique binding of segment data to sequence space to the extent that sequence numbers will not cycle through all 2**32 values before the segment data bound to those sequence numbers has been delivered and acknowledged by the receiver and all duplicate copies of the segments have "drained" from the internet. Without such an assumption, two distinct TCP segments could conceivably be

RED HAT 1010
RED HAT V. CAS

assigned the same or overlapping sequence numbers, causing confusion at the receiver as to which data is new and which is old. Remember that each segment is bound to as many consecutive sequence numbers as there are octets of data in the segment.

Under normal conditions, TCPs keep track of the next sequence number to emit and the oldest awaiting acknowledgment so as to avoid mistakenly using a sequence number over before its first use has been acknowledged. This alone does not guarantee that old duplicate data is drained from the net, so the sequence space has been made very large to reduce the probability that a wandering duplicate will cause trouble upon arrival. At 2 megabits/sec. it takes 4.5 hours to use up 2**32 octets of sequence space. Since the maximum segment lifetime in the net is not likely to exceed a few tens of seconds, this is deemed ample protection for foreseeable nets, even if data rates escalate to 10's of megabits/sec. At 100 megabits/sec, the cycle time is 5.4 minutes which may be a little short, but still within reason.

The basic duplicate detection and sequencing algorithm in TCP can be defeated, however, if a source TCP does not have any memory of the sequence numbers it last used on a given connection. For example, if the TCP were to start all connections with sequence number 0, then upon crashing and restarting, a TCP might re-form an earlier connection (possibly after half-open connection resolution) and emit packets with sequence numbers identical to or overlapping with packets still in the network which were emitted on an earlier incarnation of the same connection. In the absence of knowledge about the sequence numbers used on a particular connection, the TCP specification recommends that the source delay for MSL seconds before emitting segments on the connection, to allow time for segments from the earlier connection incarnation to drain from the system.

Even hosts which can remember the time of day and used it to select initial sequence number values are not immune from this problem (i.e., even if time of day is used to select an initial sequence number for each new connection incarnation).

Suppose, for example, that a connection is opened starting with sequence number S. Suppose that this connection is not used much and that eventually the initial sequence number function (ISN(t)) takes on a value equal to the sequence number, say S1, of the last segment sent by this TCP on a particular connection. Now suppose, at this instant, the host crashes, recovers, and establishes a new incarnation of the connection. The initial sequence number chosen is S1 = ISN(t) -- last used sequence number on old incarnation of connection! If the recovery occurs quickly enough, any old

RED HAT 1010
RED HAT V. CAS

duplicates in the net bearing sequence numbers in the neighborhood of S1 may arrive and be treated as new packets by the receiver of the new incarnation of the connection.

The problem is that the recovering host may not know for how long it crashed nor does it know whether there are still old duplicates in the system from earlier connection incarnations.

One way to deal with this problem is to deliberately delay emitting segments for one MSL after recovery from a crash- this is the "quite time" specification. Hosts which prefer to avoid waiting are willing to risk possible confusion of old and new packets at a given destination may choose not to wait for the "quite time". Implementors may provide TCP users with the ability to select on a connection by connection basis whether to wait after a crash, or may informally implement the "quite time" for all connections. Obviously, even where a user selects to "wait," this is not necessary after the host has been "up" for at least MSL seconds.

To summarize: every segment emitted occupies one or more sequence numbers in the sequence space, the numbers occupied by a segment are "busy" or "in use" until MSL seconds have passed, upon crashing a block of space-time is occupied by the octets of the last emitted segment, if a new connection is started too soon and uses any of the sequence numbers in the space-time footprint of the last segment of the previous connection incarnation, there is a potential sequence number overlap area which could cause confusion at the receiver.

3.4. Establishing a connection

The "three-way handshake" is the procedure used to establish a connection. This procedure normally is initiated by one TCP and responded to by another TCP. The procedure also works if two TCP simultaneously initiate the procedure. When simultaneous attempt occurs, each TCP receives a "SYN" segment which carries no acknowledgment after it has sent a "SYN". Of course, the arrival of an old duplicate "SYN" segment can potentially make it appear, to the recipient, that a simultaneous connection initiation is in progress. Proper use of "reset" segments can disambiguate these cases.

Several examples of connection initiation follow. Although these examples do not show connection synchronization using data-carrying segments, this is perfectly legitimate, so long as the receiving TCP doesn't deliver the data to the user until it is clear the data is valid (i.e., the data must be buffered at the receiver until the connection reaches the ESTABLISHED state). The three-way handshake reduces the possibility of false connections. It is the

RED HAT 1010
RED HAT V. CAS

implementation of a trade-off between memory and messages to provide information for this checking.

The simplest three-way handshake is shown in figure 7 below. The figures should be interpreted in the following way. Each line is numbered for reference purposes. Right arrows (-->) indicate departure of a TCP segment from TCP A to TCP B, or arrival of a segment at B from A. Left arrows (<--), indicate the reverse. Ellipsis (...) indicates a segment which is still in the network (delayed). An "XXX" indicates a segment which is lost or rejected. Comments appear in parentheses. TCP states represent the state AFTER the departure or arrival of the segment (whose contents are shown in the center of each line). Segment contents are shown in abbreviated form, with sequence number, control flags, and ACK field. Other fields such as window, addresses, lengths, and text have been left out in the interest of clarity.

```
    TCP A                                                TCP B

1.  CLOSED                                               LISTEN

2.  SYN-SENT    --> <SEQ=100><CTL=SYN>               --> SYN-RECEIVED

3.  ESTABLISHED <-- <SEQ=300><ACK=101><CTL=SYN,ACK>  <-- SYN-RECEIVED

4.  ESTABLISHED --> <SEQ=101><ACK=301><CTL=ACK>       --> ESTABLISHED

5.  ESTABLISHED --> <SEQ=101><ACK=301><CTL=ACK><DATA> --> ESTABLISHED
```

Basic 3-Way Handshake for Connection Synchronization

Figure 7.

In line 2 of figure 7, TCP A begins by sending a SYN segment indicating that it will use sequence numbers starting with sequence number 100. In line 3, TCP B sends a SYN and acknowledges the SYN it received from TCP A. Note that the acknowledgment field indicates TCP B is now expecting to hear sequence 101, acknowledging the SYN which occupied sequence 100.

At line 4, TCP A responds with an empty segment containing an ACK for TCP B's SYN; and in line 5, TCP A sends some data. Note that the sequence number of the segment in line 5 is the same as in line 4 because the ACK does not occupy sequence number space (if it did, we would wind up ACKing ACK's!).

RED HAT 1010
RED HAT V. CAS

Simultaneous initiation is only slightly more complex, as is shown in figure 8. Each TCP cycles from CLOSED to SYN-SENT to SYN-RECEIVED to ESTABLISHED.

```
    TCP A                                                TCP B

1.  CLOSED                                               CLOSED

2.  SYN-SENT     --> <SEQ=100><CTL=SYN>                  ...

3.  SYN-RECEIVED <-- <SEQ=300><CTL=SYN>                  <-- SYN-SENT

4.               ... <SEQ=100><CTL=SYN>                  --> SYN-RECEIVED

5.  SYN-RECEIVED --> <SEQ=100><ACK=301><CTL=SYN,ACK>     ...

6.  ESTABLISHED  <-- <SEQ=300><ACK=101><CTL=SYN,ACK>     <-- SYN-RECEIVED

7.               ... <SEQ=101><ACK=301><CTL=ACK>         --> ESTABLISHED
```

Simultaneous Connection Synchronization

Figure 8.

The principle reason for the three-way handshake is to prevent old duplicate connection initiations from causing confusion. To deal with this, a special control message, reset, has been devised. If the receiving TCP is in a non-synchronized state (i.e., SYN-SENT, SYN-RECEIVED), it returns to LISTEN on receiving an acceptable reset. If the TCP is in one of the synchronized states (ESTABLISHED, FIN-WAIT-1, FIN-WAIT-2, CLOSE-WAIT, CLOSING, LAST-ACK, TIME-WAIT), it aborts the connection and informs its user. We discuss this latter case under "half-open" connections below.

RED HAT 1010
RED HAT V. CAS

```
    TCP A                                                TCP B

1.  CLOSED                                               LISTEN

2.  SYN-SENT    --> <SEQ=100><CTL=SYN>               ...

3.  (duplicate) ... <SEQ=90><CTL=SYN>               --> SYN-RECEIVED

4.  SYN-SENT    <-- <SEQ=300><ACK=91><CTL=SYN,ACK>  <-- SYN-RECEIVED

5.  SYN-SENT    --> <SEQ=91><CTL=RST>               --> LISTEN

6.              ... <SEQ=100><CTL=SYN>               --> SYN-RECEIVED

7.  SYN-SENT    <-- <SEQ=400><ACK=101><CTL=SYN,ACK>  <-- SYN-RECEIVED

8.  ESTABLISHED --> <SEQ=101><ACK=401><CTL=ACK>      --> ESTABLISHED
```

Recovery from Old Duplicate SYN

Figure 9.

As a simple example of recovery from old duplicates, consider figure 9. At line 3, an old duplicate SYN arrives at TCP B. TCP B cannot tell that this is an old duplicate, so it responds normally (line 4). TCP A detects that the ACK field is incorrect and returns a RST (reset) with its SEQ field selected to make the segment believable. TCP B, on receiving the RST, returns to the LISTEN state. When the original SYN (pun intended) finally arrives at line 6, the synchronization proceeds normally. If the SYN at line 6 had arrived before the RST, a more complex exchange might have occurred with RST's sent in both directions.

Half-Open Connections and Other Anomalies

An established connection is said to be "half-open" if one of the TCPs has closed or aborted the connection at its end without the knowledge of the other, or if the two ends of the connection have become desynchronized owing to a crash that resulted in loss of memory. Such connections will automatically become reset if an attempt is made to send data in either direction. However, half-open connections are expected to be unusual, and the recovery procedure is mildly involved.

If at site A the connection no longer exists, then an attempt by the

RED HAT 1010
RED HAT V. CAS

user at site B to send any data on it will result in the site B TCP receiving a reset control message. Such a message indicates to the site B TCP that something is wrong, and it is expected to abort the connection.

Assume that two user processes A and B are communicating with one another when a crash occurs causing loss of memory to A's TCP. Depending on the operating system supporting A's TCP, it is likely that some error recovery mechanism exists. When the TCP is up again, A is likely to start again from the beginning or from a recovery point. As a result, A will probably try to OPEN the connection again or try to SEND on the connection it believes open. In the latter case, it receives the error message "connection not open" from the local (A's) TCP. In an attempt to establish the connection, A's TCP will send a segment containing SYN. This scenario leads to the example shown in figure 10. After TCP A crashes, the user attempts to re-open the connection. TCP B, in the meantime, thinks the connection is open.

```
    TCP A                                                TCP B

1.  (CRASH)                                  (send 300,receive 100)

2.  CLOSED                                               ESTABLISHED

3.  SYN-SENT --> <SEQ=400><CTL=SYN>              --> (??)

4.  (!!)     <-- <SEQ=300><ACK=100><CTL=ACK>      <-- ESTABLISHED

5.  SYN-SENT --> <SEQ=100><CTL=RST>              --> (Abort!!)

6.  SYN-SENT                                             CLOSED

7.  SYN-SENT --> <SEQ=400><CTL=SYN>              -->
```

Half-Open Connection Discovery

Figure 10.

When the SYN arrives at line 3, TCP B, being in a synchronized state, and the incoming segment outside the window, responds with an acknowledgment indicating what sequence it next expects to hear (ACK 100). TCP A sees that this segment does not acknowledge anything it sent and, being unsynchronized, sends a reset (RST) because it has detected a half-open connection. TCP B aborts at line 5. TCP A will

RED HAT 1010
RED HAT V. CAS

continue to try to establish the connection; the problem is now reduced to the basic 3-way handshake of figure 7.

An interesting alternative case occurs when TCP A crashes and TCP B tries to send data on what it thinks is a synchronized connection. This is illustrated in figure 11. In this case, the data arriving at TCP A from TCP B (line 2) is unacceptable because no such connection exists, so TCP A sends a RST. The RST is acceptable so TCP B processes it and aborts the connection.

```
      TCP A                                                TCP B

1.  (CRASH)                                      (send 300,receive 100)

2.  (??)    <-- <SEQ=300><ACK=100><DATA=10><CTL=ACK> <-- ESTABLISHED

3.          --> <SEQ=100><CTL=RST>                   --> (ABORT!!)
```

Active Side Causes Half-Open Connection Discovery

Figure 11.

In figure 12, we find the two TCPs A and B with passive connections waiting for SYN. An old duplicate arriving at TCP B (line 2) stirs B into action. A SYN-ACK is returned (line 3) and causes TCP A to generate a RST (the ACK in line 3 is not acceptable). TCP B accepts the reset and returns to its passive LISTEN state.

```
    TCP A                                          TCP B

1.  LISTEN                                         LISTEN

2.       ... <SEQ=Z><CTL=SYN>                -->  SYN-RECEIVED

3.  (??) <-- <SEQ=X><ACK=Z+1><CTL=SYN,ACK>   <--  SYN-RECEIVED

4.       --> <SEQ=Z+1><CTL=RST>              -->  (return to LISTEN!)

5.  LISTEN                                         LISTEN
```

Old Duplicate SYN Initiates a Reset on two Passive Sockets

Figure 12.

RED HAT 1010
RED HAT V. CAS

A variety of other cases are possible, all of which are accounted for by the following rules for RST generation and processing.

Reset Generation

As a general rule, reset (RST) must be sent whenever a segment arrives which apparently is not intended for the current connection. A reset must not be sent if it is not clear that this is the case.

There are three groups of states:

1. If the connection does not exist (CLOSED) then a reset is sent in response to any incoming segment except another reset. In particular, SYNs addressed to a non-existent connection are rejected by this means.

   If the incoming segment has an ACK field, the reset takes its sequence number from the ACK field of the segment, otherwise the reset has sequence number zero and the ACK field is set to the sum of the sequence number and segment length of the incoming segment. The connection remains in the CLOSED state.

2. If the connection is in any non-synchronized state (LISTEN, SYN-SENT, SYN-RECEIVED), and the incoming segment acknowledges something not yet sent (the segment carries an unacceptable ACK), or if an incoming segment has a security level or compartment which does not exactly match the level and compartment requested for the connection, a reset is sent.

   If our SYN has not been acknowledged and the precedence level of the incoming segment is higher than the precedence level requested then either raise the local precedence level (if allowed by the user and the system) or send a reset; or if the precedence level of the incoming segment is lower than the precedence level requested then continue as if the precedence matched exactly (if the remote TCP cannot raise the precedence level to match ours this will be detected in the next segment it sends, and the connection will be terminated then). If our SYN has been acknowledged (perhaps in this incoming segment) the precedence level of the incoming segment must match the local precedence level exactly, if it does not a reset must be sent.

   If the incoming segment has an ACK field, the reset takes its sequence number from the ACK field of the segment, otherwise the reset has sequence number zero and the ACK field is set to the sum of the sequence number and segment length of the incoming segment. The connection remains in the same state.

RED HAT 1010
RED HAT V. CAS

3. If the connection is in a synchronized state (ESTABLISHED, FIN-WAIT-1, FIN-WAIT-2, CLOSE-WAIT, CLOSING, LAST-ACK, TIME-WAIT), any unacceptable segment (out of window sequence number or unacceptible acknowledgment number) must elicit only an empty acknowledgment segment containing the current send-sequence number and an acknowledgment indicating the next sequence number expected to be received, and the connection remains in the same state.

If an incoming segment has a security level, or compartment, or precedence which does not exactly match the level, and compartment, and precedence requested for the connection,a reset is sent and connection goes to the CLOSED state. The reset takes its sequence number from the ACK field of the incoming segment.

Reset Processing

In all states except SYN-SENT, all reset (RST) segments are validated by checking their SEQ-fields. A reset is valid if its sequence number is in the window. In the SYN-SENT state (a RST received in response to an initial SYN), the RST is acceptable if the ACK field acknowledges the SYN.

The receiver of a RST first validates it, then changes state. If the receiver was in the LISTEN state, it ignores it. If the receiver was in SYN-RECEIVED state and had previously been in the LISTEN state, then the receiver returns to the LISTEN state, otherwise the receiver aborts the connection and goes to the CLOSED state. If the receiver was in any other state, it aborts the connection and advises the user and goes to the CLOSED state.

3.5. Closing a Connection

CLOSE is an operation meaning "I have no more data to send." The notion of closing a full-duplex connection is subject to ambiguous interpretation, of course, since it may not be obvious how to treat the receiving side of the connection. We have chosen to treat CLOSE in a simplex fashion. The user who CLOSEs may continue to RECEIVE until he is told that the other side has CLOSED also. Thus, a program could initiate several SENDs followed by a CLOSE, and then continue to RECEIVE until signaled that a RECEIVE failed because the other side has CLOSED. We assume that the TCP will signal a user, even if no RECEIVEs are outstanding, that the other side has closed, so the user can terminate his side gracefully. A TCP will reliably deliver all buffers SENT before the connection was CLOSED so a user who expects no data in return need only wait to hear the connection was CLOSED successfully to know that all his data was received at the destination TCP. Users must keep reading connections they close for sending until the TCP says no more data.

RED HAT 1010
RED HAT V. CAS

There are essentially three cases:

1) The user initiates by telling the TCP to CLOSE the connection

2) The remote TCP initiates by sending a FIN control signal

3) Both users CLOSE simultaneously

Case 1: Local user initiates the close

In this case, a FIN segment can be constructed and placed on the outgoing segment queue. No further SENDs from the user will be accepted by the TCP, and it enters the FIN-WAIT-1 state. RECEIVEs are allowed in this state. All segments preceding and including FIN will be retransmitted until acknowledged. When the other TCP has both acknowledged the FIN and sent a FIN of its own, the first TCP can ACK this FIN. Note that a TCP receiving a FIN will ACK but not send its own FIN until its user has CLOSED the connection also.

Case 2: TCP receives a FIN from the network

If an unsolicited FIN arrives from the network, the receiving TCP can ACK it and tell the user that the connection is closing. The user will respond with a CLOSE, upon which the TCP can send a FIN to the other TCP after sending any remaining data. The TCP then waits until its own FIN is acknowledged whereupon it deletes the connection. If an ACK is not forthcoming, after the user timeout the connection is aborted and the user is told.

Case 3: both users close simultaneously

A simultaneous CLOSE by users at both ends of a connection causes FIN segments to be exchanged. When all segments preceding the FINs have been processed and acknowledged, each TCP can ACK the FIN it has received. Both will, upon receiving these ACKs, delete the connection.

```
    TCP A                                                  TCP B

1.  ESTABLISHED                                            ESTABLISHED

2.  (Close)
    FIN-WAIT-1  --> <SEQ=100><ACK=300><CTL=FIN,ACK>  --> CLOSE-WAIT

3.  FIN-WAIT-2  <-- <SEQ=300><ACK=101><CTL=ACK>      <-- CLOSE-WAIT

4.                                                         (Close)
    TIME-WAIT   <-- <SEQ=300><ACK=101><CTL=FIN,ACK>  <-- LAST-ACK

5.  TIME-WAIT   --> <SEQ=101><ACK=301><CTL=ACK>      --> CLOSED

6.  (2 MSL)
    CLOSED
```

Normal Close Sequence

Figure 13.

```
    TCP A                                                  TCP B

1.  ESTABLISHED                                            ESTABLISHED

2.  (Close)                                                (Close)
    FIN-WAIT-1  --> <SEQ=100><ACK=300><CTL=FIN,ACK>  ... FIN-WAIT-1
                <-- <SEQ=300><ACK=100><CTL=FIN,ACK>  <--
                ... <SEQ=100><ACK=300><CTL=FIN,ACK>  -->

3.  CLOSING     --> <SEQ=101><ACK=301><CTL=ACK>      ... CLOSING
                <-- <SEQ=301><ACK=101><CTL=ACK>      <--
                ... <SEQ=101><ACK=301><CTL=ACK>      -->

4.  TIME-WAIT                                              TIME-WAIT
    (2 MSL)                                                (2 MSL)
    CLOSED                                                 CLOSED
```

Simultaneous Close Sequence

Figure 14.

RED HAT 1010
RED HAT V. CAS

3.6. Precedence and Security

The intent is that connection be allowed only between ports operating with exactly the same security and compartment values and at the higher of the precedence level requested by the two ports.

The precedence and security parameters used in TCP are exactly those defined in the Internet Protocol (IP) [2]. Throughout this TCP specification the term "security/compartment" is intended to indicate the security parameters used in IP including security, compartment, user group, and handling restriction.

A connection attempt with mismatched security/compartment values or a lower precedence value must be rejected by sending a reset. Rejecting a connection due to too low a precedence only occurs after an acknowledgment of the SYN has been received.

Note that TCP modules which operate only at the default value of precedence will still have to check the precedence of incoming segments and possibly raise the precedence level they use on the connection.

The security paramaters may be used even in a non-secure environment (the values would indicate unclassified data), thus hosts in non-secure environments must be prepared to receive the security parameters, though they need not send them.

3.7. Data Communication

Once the connection is established data is communicated by the exchange of segments. Because segments may be lost due to errors (checksum test failure), or network congestion, TCP uses retransmission (after a timeout) to ensure delivery of every segment. Duplicate segments may arrive due to network or TCP retransmission. As discussed in the section on sequence numbers the TCP performs certain tests on the sequence and acknowledgment numbers in the segments to verify their acceptability.

The sender of data keeps track of the next sequence number to use in the variable SND.NXT. The receiver of data keeps track of the next sequence number to expect in the variable RCV.NXT. The sender of data keeps track of the oldest unacknowledged sequence number in the variable SND.UNA. If the data flow is momentarily idle and all data sent has been acknowledged then the three variables will be equal.

When the sender creates a segment and transmits it the sender advances SND.NXT. When the receiver accepts a segment it advances RCV.NXT and sends an acknowledgment. When the data sender receives an

RED HAT 1010
RED HAT V. CAS

acknowledgment it advances SND.UNA. The extent to which the values of these variables differ is a measure of the delay in the communication. The amount by which the variables are advanced is the length of the data in the segment. Note that once in the ESTABLISHED state all segments must carry current acknowledgment information.

The CLOSE user call implies a push function, as does the FIN control flag in an incoming segment.

Retransmission Timeout

Because of the variability of the networks that compose an internetwork system and the wide range of uses of TCP connections the retransmission timeout must be dynamically determined. One procedure for determining a retransmission time out is given here as an illustration.

An Example Retransmission Timeout Procedure

Measure the elapsed time between sending a data octet with a particular sequence number and receiving an acknowledgment that covers that sequence number (segments sent do not have to match segments received). This measured elapsed time is the Round Trip Time (RTT). Next compute a Smoothed Round Trip Time (SRTT) as:

SRTT = (ALPHA * SRTT) + ((1-ALPHA) * RTT)

and based on this, compute the retransmission timeout (RTO) as:

RTO = min[UBOUND,max[LBOUND,(BETA*SRTT)]]

where UBOUND is an upper bound on the timeout (e.g., 1 minute), LBOUND is a lower bound on the timeout (e.g., 1 second), ALPHA is a smoothing factor (e.g., .8 to .9), and BETA is a delay variance factor (e.g., 1.3 to 2.0).

The Communication of Urgent Information

The objective of the TCP urgent mechanism is to allow the sending user to stimulate the receiving user to accept some urgent data and to permit the receiving TCP to indicate to the receiving user when all the currently known urgent data has been received by the user.

This mechanism permits a point in the data stream to be designated as the end of urgent information. Whenever this point is in advance of the receive sequence number (RCV.NXT) at the receiving TCP, that TCP must tell the user to go into "urgent mode"; when the receive sequence number catches up to the urgent pointer, the TCP must tell user to go

RED HAT 1010
RED HAT V. CAS

into "normal mode". If the urgent pointer is updated while the user is in "urgent mode", the update will be invisible to the user.

The method employs a urgent field which is carried in all segments transmitted. The URG control flag indicates that the urgent field is meaningful and must be added to the segment sequence number to yield the urgent pointer. The absence of this flag indicates that there is no urgent data outstanding.

To send an urgent indication the user must also send at least one data octet. If the sending user also indicates a push, timely delivery of the urgent information to the destination process is enhanced.

Managing the Window

The window sent in each segment indicates the range of sequence numbers the sender of the window (the data receiver) is currently prepared to accept. There is an assumption that this is related to the currently available data buffer space available for this connection.

Indicating a large window encourages transmissions. If more data arrives than can be accepted, it will be discarded. This will result in excessive retransmissions, adding unnecessarily to the load on the network and the TCPs. Indicating a small window may restrict the transmission of data to the point of introducing a round trip delay between each new segment transmitted.

The mechanisms provided allow a TCP to advertise a large window and to subsequently advertise a much smaller window without having accepted that much data. This, so called "shrinking the window," is strongly discouraged. The robustness principle dictates that TCPs will not shrink the window themselves, but will be prepared for such behavior on the part of other TCPs.

The sending TCP must be prepared to accept from the user and send at least one octet of new data even if the send window is zero. The sending TCP must regularly retransmit to the receiving TCP even when the window is zero. Two minutes is recommended for the retransmission interval when the window is zero. This retransmission is essential to guarantee that when either TCP has a zero window the re-opening of the window will be reliably reported to the other.

When the receiving TCP has a zero window and a segment arrives it must still send an acknowledgment showing its next expected sequence number and current window (zero).

The sending TCP packages the data to be transmitted into segments

RED HAT 1010
RED HAT V. CAS

which fit the current window, and may repackage segments on the retransmission queue. Such repackaging is not required, but may be helpful.

In a connection with a one-way data flow, the window information will be carried in acknowledgment segments that all have the same sequence number so there will be no way to reorder them if they arrive out of order. This is not a serious problem, but it will allow the window information to be on occasion temporarily based on old reports from the data receiver. A refinement to avoid this problem is to act on the window information from segments that carry the highest acknowledgment number (that is segments with acknowledgment number equal or greater than the highest previously received).

The window management procedure has significant influence on the communication performance. The following comments are suggestions to implementers.

Window Management Suggestions

Allocating a very small window causes data to be transmitted in many small segments when better performance is achieved using fewer large segments.

One suggestion for avoiding small windows is for the receiver to defer updating a window until the additional allocation is at least X percent of the maximum allocation possible for the connection (where X might be 20 to 40).

Another suggestion is for the sender to avoid sending small segments by waiting until the window is large enough before sending data. If the the user signals a push function then the data must be sent even if it is a small segment.

Note that the acknowledgments should not be delayed or unnecessary retransmissions will result. One strategy would be to send an acknowledgment when a small segment arrives (with out updating the window information), and then to send another acknowledgment with new window information when the window is larger.

The segment sent to probe a zero window may also begin a break up of transmitted data into smaller and smaller segments. If a segment containing a single data octet sent to probe a zero window is accepted, it consumes one octet of the window now available. If the sending TCP simply sends as much as it can whenever the window is non zero, the transmitted data will be broken into alternating big and small segments. As time goes on, occasional pauses in the receiver making window allocation available will

RED HAT 1010
RED HAT V. CAS

result in breaking the big segments into a small and not quite so big pair. And after a while the data transmission will be in mostly small segments.

The suggestion here is that the TCP implementations need to actively attempt to combine small window allocations into larger windows, since the mechanisms for managing the window tend to lead to many small windows in the simplest minded implementations.

3.8. Interfaces

There are of course two interfaces of concern: the user/TCP interface and the TCP/lower-level interface. We have a fairly elaborate model of the user/TCP interface, but the interface to the lower level protocol module is left unspecified here, since it will be specified in detail by the specification of the lowel level protocol. For the case that the lower level is IP we note some of the parameter values that TCPs might use.

User/TCP Interface

The following functional description of user commands to the TCP is, at best, fictional, since every operating system will have different facilities. Consequently, we must warn readers that different TCP implementations may have different user interfaces. However, all TCPs must provide a certain minimum set of services to guarantee that all TCP implementations can support the same protocol hierarchy. This section specifies the functional interfaces required of all TCP implementations.

TCP User Commands

The following sections functionally characterize a USER/TCP interface. The notation used is similar to most procedure or function calls in high level languages, but this usage is not meant to rule out trap type service calls (e.g., SVCs, UUOs, EMTs).

The user commands described below specify the basic functions the TCP must perform to support interprocess communication. Individual implementations must define their own exact format, and may provide combinations or subsets of the basic functions in single calls. In particular, some implementations may wish to automatically OPEN a connection on the first SEND or RECEIVE issued by the user for a given connection.

RED HAT 1010
RED HAT V. CAS

In providing interprocess communication facilities, the TCP must not only accept commands, but must also return information to the processes it serves. The latter consists of:

(a) general information about a connection (e.g., interrupts, remote close, binding of unspecified foreign socket).

(b) replies to specific user commands indicating success or various types of failure.

Open

Format: OPEN (local port, foreign socket, active/passive [, timeout] [, precedence] [, security/compartment] [, options]) -> local connection name

We assume that the local TCP is aware of the identity of the processes it serves and will check the authority of the process to use the connection specified. Depending upon the implementation of the TCP, the local network and TCP identifiers for the source address will either be supplied by the TCP or the lower level protocol (e.g., IP). These considerations are the result of concern about security, to the extent that no TCP be able to masquerade as another one, and so on. Similarly, no process can masquerade as another without the collusion of the TCP.

If the active/passive flag is set to passive, then this is a call to LISTEN for an incoming connection. A passive open may have either a fully specified foreign socket to wait for a particular connection or an unspecified foreign socket to wait for any call. A fully specified passive call can be made active by the subsequent execution of a SEND.

A transmission control block (TCB) is created and partially filled in with data from the OPEN command parameters.

On an active OPEN command, the TCP will begin the procedure to synchronize (i.e., establish) the connection at once.

The timeout, if present, permits the caller to set up a timeout for all data submitted to TCP. If data is not successfully delivered to the destination within the timeout period, the TCP will abort the connection. The present global default is five minutes.

The TCP or some component of the operating system will verify the users authority to open a connection with the specified

RED HAT 1010
RED HAT V. CAS

precedence or security/compartment. The absence of precedence or security/compartment specification in the OPEN call indicates the default values must be used.

TCP will accept incoming requests as matching only if the security/compartment information is exactly the same and only if the precedence is equal to or higher than the precedence requested in the OPEN call.

The precedence for the connection is the higher of the values requested in the OPEN call and received from the incoming request, and fixed at that value for the life of the connection.Implementers may want to give the user control of this precedence negotiation. For example, the user might be allowed to specify that the precedence must be exactly matched, or that any attempt to raise the precedence be confirmed by the user.

A local connection name will be returned to the user by the TCP. The local connection name can then be used as a short hand term for the connection defined by the <local socket, foreign socket> pair.

Send

Format: SEND (local connection name, buffer address, byte count, PUSH flag, URGENT flag [,timeout])

This call causes the data contained in the indicated user buffer to be sent on the indicated connection. If the connection has not been opened, the SEND is considered an error. Some implementations may allow users to SEND first; in which case, an automatic OPEN would be done. If the calling process is not authorized to use this connection, an error is returned.

If the PUSH flag is set, the data must be transmitted promptly to the receiver, and the PUSH bit will be set in the last TCP segment created from the buffer. If the PUSH flag is not set, the data may be combined with data from subsequent SENDs for transmission efficiency.

If the URGENT flag is set, segments sent to the destination TCP will have the urgent pointer set. The receiving TCP will signal the urgent condition to the receiving process if the urgent pointer indicates that data preceding the urgent pointer has not been consumed by the receiving process. The purpose of urgent is to stimulate the receiver to process the urgent data and to indicate to the receiver when all the currently known urgent

RED HAT 1010
RED HAT V. CAS

data has been received. The number of times the sending user's TCP signals urgent will not necessarily be equal to the number of times the receiving user will be notified of the presence of urgent data.

If no foreign socket was specified in the OPEN, but the connection is established (e.g., because a LISTENing connection has become specific due to a foreign segment arriving for the local socket), then the designated buffer is sent to the implied foreign socket. Users who make use of OPEN with an unspecified foreign socket can make use of SEND without ever explicitly knowing the foreign socket address.

However, if a SEND is attempted before the foreign socket becomes specified, an error will be returned. Users can use the STATUS call to determine the status of the connection. In some implementations the TCP may notify the user when an unspecified socket is bound.

If a timeout is specified, the current user timeout for this connection is changed to the new one.

In the simplest implementation, SEND would not return control to the sending process until either the transmission was complete or the timeout had been exceeded. However, this simple method is both subject to deadlocks (for example, both sides of the connection might try to do SENDs before doing any RECEIVEs) and offers poor performance, so it is not recommended. A more sophisticated implementation would return immediately to allow the process to run concurrently with network I/O, and, furthermore, to allow multiple SENDs to be in progress. Multiple SENDs are served in first come, first served order, so the TCP will queue those it cannot service immediately.

We have implicitly assumed an asynchronous user interface in which a SEND later elicits some kind of SIGNAL or pseudo-interrupt from the serving TCP. An alternative is to return a response immediately. For instance, SENDs might return immediate local acknowledgment, even if the segment sent had not been acknowledged by the distant TCP. We could optimistically assume eventual success. If we are wrong, the connection will close anyway due to the timeout. In implementations of this kind (synchronous), there will still be some asynchronous signals, but these will deal with the connection itself, and not with specific segments or buffers.

In order for the process to distinguish among error or success indications for different SENDs, it might be appropriate for the

RED HAT 1010
RED HAT V. CAS

buffer address to be returned along with the coded response to the SEND request. TCP-to-user signals are discussed below, indicating the information which should be returned to the calling process.

Receive

Format: RECEIVE (local connection name, buffer address, byte count) -> byte count, urgent flag, push flag

This command allocates a receiving buffer associated with the specified connection. If no OPEN precedes this command or the calling process is not authorized to use this connection, an error is returned.

In the simplest implementation, control would not return to the calling program until either the buffer was filled, or some error occurred, but this scheme is highly subject to deadlocks. A more sophisticated implementation would permit several RECEIVEs to be outstanding at once. These would be filled as segments arrive. This strategy permits increased throughput at the cost of a more elaborate scheme (possibly asynchronous) to notify the calling program that a PUSH has been seen or a buffer filled.

If enough data arrive to fill the buffer before a PUSH is seen, the PUSH flag will not be set in the response to the RECEIVE. The buffer will be filled with as much data as it can hold. If a PUSH is seen before the buffer is filled the buffer will be returned partially filled and PUSH indicated.

If there is urgent data the user will have been informed as soon as it arrived via a TCP-to-user signal. The receiving user should thus be in "urgent mode". If the URGENT flag is on, additional urgent data remains. If the URGENT flag is off, this call to RECEIVE has returned all the urgent data, and the user may now leave "urgent mode". Note that data following the urgent pointer (non-urgent data) cannot be delivered to the user in the same buffer with preceeding urgent data unless the boundary is clearly marked for the user.

To distinguish among several outstanding RECEIVEs and to take care of the case that a buffer is not completely filled, the return code is accompanied by both a buffer pointer and a byte count indicating the actual length of the data received.

Alternative implementations of RECEIVE might have the TCP

RED HAT 1010
RED HAT V. CAS

allocate buffer storage, or the TCP might share a ring buffer with the user.

Close

Format: CLOSE (local connection name)

This command causes the connection specified to be closed. If the connection is not open or the calling process is not authorized to use this connection, an error is returned. Closing connections is intended to be a graceful operation in the sense that outstanding SENDs will be transmitted (and retransmitted), as flow control permits, until all have been serviced. Thus, it should be acceptable to make several SEND calls, followed by a CLOSE, and expect all the data to be sent to the destination. It should also be clear that users should continue to RECEIVE on CLOSING connections, since the other side may be trying to transmit the last of its data. Thus, CLOSE means "I have no more to send" but does not mean "I will not receive any more." It may happen (if the user level protocol is not well thought out) that the closing side is unable to get rid of all its data before timing out. In this event, CLOSE turns into ABORT, and the closing TCP gives up.

The user may CLOSE the connection at any time on his own initiative, or in response to various prompts from the TCP (e.g., remote close executed, transmission timeout exceeded, destination inaccessible).

Because closing a connection requires communication with the foreign TCP, connections may remain in the closing state for a short time. Attempts to reopen the connection before the TCP replies to the CLOSE command will result in error responses.

Close also implies push function.

Status

Format: STATUS (local connection name) -> status data

This is an implementation dependent user command and could be excluded without adverse effect. Information returned would typically come from the TCB associated with the connection.

This command returns a data block containing the following information:

local socket,

foreign socket,
local connection name,
receive window,
send window,
connection state,
number of buffers awaiting acknowledgment,
number of buffers pending receipt,
urgent state,
precedence,
security/compartment,
and transmission timeout.

Depending on the state of the connection, or on the implementation itself, some of this information may not be available or meaningful. If the calling process is not authorized to use this connection, an error is returned. This prevents unauthorized processes from gaining information about a connection.

Abort

Format: ABORT (local connection name)

This command causes all pending SENDs and RECEIVES to be aborted, the TCB to be removed, and a special RESET message to be sent to the TCP on the other side of the connection. Depending on the implementation, users may receive abort indications for each outstanding SEND or RECEIVE, or may simply receive an ABORT-acknowledgment.

TCP-to-User Messages

It is assumed that the operating system environment provides a means for the TCP to asynchronously signal the user program. When the TCP does signal a user program, certain information is passed to the user. Often in the specification the information will be an error message. In other cases there will be information relating to the completion of processing a SEND or RECEIVE or other user call.

The following information is provided:

Local Connection Name	Always
Response String	Always
Buffer Address	Send & Receive
Byte count (counts bytes received)	Receive
Push flag	Receive
Urgent flag	Receive

TCP/Lower-Level Interface

The TCP calls on a lower level protocol module to actually send and receive information over a network. One case is that of the ARPA internetwork system where the lower level module is the Internet Protocol (IP) [2].

If the lower level protocol is IP it provides arguments for a type of service and for a time to live. TCP uses the following settings for these parameters:

Type of Service = Precedence: routine, Delay: normal, Throughput: normal, Reliability: normal; or 00000000.

Time to Live = one minute, or 00111100.

Note that the assumed maximum segment lifetime is two minutes. Here we explicitly ask that a segment be destroyed if it cannot be delivered by the internet system within one minute.

If the lower level is IP (or other protocol that provides this feature) and source routing is used, the interface must allow the route information to be communicated. This is especially important so that the source and destination addresses used in the TCP checksum be the originating source and ultimate destination. It is also important to preserve the return route to answer connection requests.

Any lower level protocol will have to provide the source address, destination address, and protocol fields, and some way to determine the "TCP length", both to provide the functional equivlent service of IP and to be used in the TCP checksum.

RED HAT 1010
RED HAT V. CAS

3.9. Event Processing

The processing depicted in this section is an example of one possible implementation. Other implementations may have slightly different processing sequences, but they should differ from those in this section only in detail, not in substance.

The activity of the TCP can be characterized as responding to events. The events that occur can be cast into three categories: user calls, arriving segments, and timeouts. This section describes the processing the TCP does in response to each of the events. In many cases the processing required depends on the state of the connection.

Events that occur:

User Calls

OPEN
SEND
RECEIVE
CLOSE
ABORT
STATUS

Arriving Segments

SEGMENT ARRIVES

Timeouts

USER TIMEOUT
RETRANSMISSION TIMEOUT
TIME-WAIT TIMEOUT

The model of the TCP/user interface is that user commands receive an immediate return and possibly a delayed response via an event or pseudo interrupt. In the following descriptions, the term "signal" means cause a delayed response.

Error responses are given as character strings. For example, user commands referencing connections that do not exist receive "error: connection not open".

Please note in the following that all arithmetic on sequence numbers, acknowledgment numbers, windows, et cetera, is modulo 2**32 the size of the sequence number space. Also note that "=<" means less than or equal to (modulo 2**32).

RED HAT 1010
RED HAT V. CAS

A natural way to think about processing incoming segments is to imagine that they are first tested for proper sequence number (i.e., that their contents lie in the range of the expected "receive window" in the sequence number space) and then that they are generally queued and processed in sequence number order.

When a segment overlaps other already received segments we reconstruct the segment to contain just the new data, and adjust the header fields to be consistent.

Note that if no state change is mentioned the TCP stays in the same state.

RED HAT 1010
RED HAT V. CAS

OPEN Call

CLOSED STATE (i.e., TCB does not exist)

Create a new transmission control block (TCB) to hold connection state information. Fill in local socket identifier, foreign socket, precedence, security/compartment, and user timeout information. Note that some parts of the foreign socket may be unspecified in a passive OPEN and are to be filled in by the parameters of the incoming SYN segment. Verify the security and precedence requested are allowed for this user, if not return "error: precedence not allowed" or "error: security/compartment not allowed." If passive enter the LISTEN state and return. If active and the foreign socket is unspecified, return "error: foreign socket unspecified"; if active and the foreign socket is specified, issue a SYN segment. An initial send sequence number (ISS) is selected. A SYN segment of the form <SEQ=ISS><CTL=SYN> is sent. Set SND.UNA to ISS, SND.NXT to ISS+1, enter SYN-SENT state, and return.

If the caller does not have access to the local socket specified, return "error: connection illegal for this process". If there is no room to create a new connection, return "error: insufficient resources".

LISTEN STATE

If active and the foreign socket is specified, then change the connection from passive to active, select an ISS. Send a SYN segment, set SND.UNA to ISS, SND.NXT to ISS+1. Enter SYN-SENT state. Data associated with SEND may be sent with SYN segment or queued for transmission after entering ESTABLI$HED state. The urgent bit if requested in the command must be sent with the data segments sent as a result of this command. If there is no room to queue the request, respond with "error: insufficient resources". If Foreign socket was not specified, then return "error: foreign socket unspecified".

OPEN Call

SYN-SENT STATE
SYN-RECEIVED STATE
ESTABLISHED STATE
FIN-WAIT-1 STATE
FIN-WAIT-2 STATE
CLOSE-WAIT STATE
CLOSING STATE
LAST-ACK STATE
TIME-WAIT STATE

Return "error: connection already exists".

SEND Call

CLOSED STATE (i.e., TCB does not exist)

If the user does not have access to such a connection, then return "error: connection illegal for this process".

Otherwise, return "error: connection does not exist".

LISTEN STATE

If the foreign socket is specified, then change the connection from passive to active, select an ISS. Send a SYN segment, set SND.UNA to ISS, SND.NXT to ISS+1. Enter SYN-SENT state. Data associated with SEND may be sent with SYN segment or queued for transmission after entering ESTABLISHED state. The urgent bit if requested in the command must be sent with the data segments sent as a result of this command. If there is no room to queue the request, respond with "error: insufficient resources". If Foreign socket was not specified, then return "error: foreign socket unspecified".

SYN-SENT STATE
SYN-RECEIVED STATE

Queue the data for transmission after entering ESTABLISHED state. If no space to queue, respond with "error: insufficient resources".

ESTABLISHED STATE
CLOSE-WAIT STATE

Segmentize the buffer and send it with a piggybacked acknowledgment (acknowledgment value = RCV.NXT). If there is insufficient space to remember this buffer, simply return "error: insufficient resources".

If the urgent flag is set, then SND.UP <- SND.NXT-1 and set the urgent pointer in the outgoing segments.

RED HAT 1010
RED HAT V. CAS

SEND Call

FIN-WAIT-1 STATE
FIN-WAIT-2 STATE
CLOSING STATE
LAST-ACK STATE
TIME-WAIT STATE

Return "error: connection closing" and do not service request.

RECEIVE Call

CLOSED STATE (i.e., TCB does not exist)

If the user does not have access to such a connection, return "error: connection illegal for this process".

Otherwise return "error: connection does not exist".

LISTEN STATE
SYN-SENT STATE
SYN-RECEIVED STATE

Queue for processing after entering ESTABLISHED state. If there is no room to queue this request, respond with "error: insufficient resources".

ESTABLISHED STATE
FIN-WAIT-1 STATE
FIN-WAIT-2 STATE

If insufficient incoming segments are queued to satisfy the request, queue the request. If there is no queue space to remember the RECEIVE, respond with "error: insufficient resources".

Reassemble queued incoming segments into receive buffer and return to user. Mark "push seen" (PUSH) if this is the case.

If RCV.UP is in advance of the data currently being passed to the user notify the user of the presence of urgent data.

When the TCP takes responsibility for delivering data to the user that fact must be communicated to the sender via an acknowledgment. The formation of such an acknowledgment is described below in the discussion of processing an incoming segment.

CLOSE-WAIT STATE

Since the remote side has already sent FIN, RECEIVEs must be satisfied by text already on hand, but not yet delivered to the user. If no text is awaiting delivery, the RECEIVE will get a "error: connection closing" response. Otherwise, any remaining text can be used to satisfy the RECEIVE.

CLOSING STATE
LAST-ACK STATE
TIME-WAIT STATE

Return "error: connection closing".

CLOSE Call

CLOSED STATE (i.e., TCB does not exist)

If the user does not have access to such a connection, return "error: connection illegal for this process".

Otherwise, return "error: connection does not exist".

LISTEN STATE

Any outstanding RECEIVEs are returned with "error: closing" responses. Delete TCB, enter CLOSED state, and return.

SYN-SENT STATE

Delete the TCB and return "error: closing" responses to any queued SENDs, or RECEIVEs.

SYN-RECEIVED STATE

If no SENDs have been issued and there is no pending data to send, then form a FIN segment and send it, and enter FIN-WAIT-1 state; otherwise queue for processing after entering ESTABLISHED state.

ESTABLISHED STATE

Queue this until all preceding SENDs have been segmentized, then form a FIN segment and send it. In any case, enter FIN-WAIT-1 state.

FIN-WAIT-1 STATE
FIN-WAIT-2 STATE

Strictly speaking, this is an error and should receive a "error: connection closing" response. An "ok" response would be acceptable, too, as long as a second FIN is not emitted (the first FIN may be retransmitted though).

CLOSE Call

CLOSE-WAIT STATE

Queue this request until all preceding SENDs have been segmentized; then send a FIN segment, enter CLOSING state.

CLOSING STATE
LAST-ACK STATE
TIME-WAIT STATE

Respond with "error: connection closing".

RED HAT 1010
RED HAT V. CAS

ABORT Call

CLOSED STATE (i.e., TCB does not exist)

If the user should not have access to such a connection, return "error: connection illegal for this process".

Otherwise return "error: connection does not exist".

LISTEN STATE

Any outstanding RECEIVEs should be returned with "error: connection reset" responses. Delete TCB, enter CLOSED state, and return.

SYN-SENT STATE

All queued SENDs and RECEIVEs should be given "connection reset" notification, delete the TCB, enter CLOSED state, and return.

SYN-RECEIVED STATE
ESTABLISHED STATE
FIN-WAIT-1 STATE
FIN-WAIT-2 STATE
CLOSE-WAIT STATE

Send a reset segment:

<SEQ=SND.NXT><CTL=RST>

All queued SENDs and RECEIVEs should be given "connection reset" notification; all segments queued for transmission (except for the RST formed above) or retransmission should be flushed, delete the TCB, enter CLOSED state, and return.

CLOSING STATE
LAST-ACK STATE
TIME-WAIT STATE

Respond with "ok" and delete the TCB, enter CLOSED state, and return.

STATUS Call

CLOSED STATE (i.e., TCB does not exist)

If the user should not have access to such a connection, return "error: connection illegal for this process".

Otherwise return "error: connection does not exist".

LISTEN STATE

Return "state = LISTEN", and the TCB pointer.

SYN-SENT STATE

Return "state = SYN-SENT", and the TCB pointer.

SYN-RECEIVED STATE

Return "state = SYN-RECEIVED", and the TCB pointer.

ESTABLISHED STATE

Return "state = ESTABLISHED", and the TCB pointer.

FIN-WAIT-1 STATE

Return "state = FIN-WAIT-1", and the TCB pointer.

FIN-WAIT-2 STATE

Return "state = FIN-WAIT-2", and the TCB pointer.

CLOSE-WAIT STATE

Return "state = CLOSE-WAIT", and the TCB pointer.

CLOSING STATE

Return "state = CLOSING", and the TCB pointer.

LAST-ACK STATE

Return "state = LAST-ACK", and the TCB pointer.

RED HAT 1010
RED HAT V. CAS

TIME-WAIT STATE

Return "state = TIME-WAIT", and the TCB pointer.

SEGMENT ARRIVES

If the state is CLOSED (i.e., TCB does not exist) then

all data in the incoming segment is discarded. An incoming segment containing a RST is discarded. An incoming segment not containing a RST causes a RST to be sent in response. The acknowledgment and sequence field values are selected to make the reset sequence acceptable to the TCP that sent the offending segment.

If the ACK bit is off, sequence number zero is used,

<SEQ=0><ACK=SEG.SEQ+SEG.LEN><CTL=RST,ACK>

If the ACK bit is on,

<SEQ=SEG.ACK><CTL=RST>

Return.

If the state is LISTEN then

first check for an RST

An incoming RST should be ignored. Return.

second check for an ACK

Any acknowledgment is bad if it arrives on a connection still in the LISTEN state. An acceptable reset segment should be formed for any arriving ACK-bearing segment. The RST should be formatted as follows:

<SEQ=SEG.ACK><CTL=RST>

Return.

third check for a SYN

If the SYN bit is set, check the security. If the security/compartment on the incoming segment does not exactly match the security/compartment in the TCB then send a reset and return.

<SEQ=SEG.ACK><CTL=RST>

RED HAT 1010
RED HAT V. CAS

If the SEG.PRC is greater than the TCB.PRC then if allowed by the user and the system set TCB.PRC<-SEG.PRC, if not allowed send a reset and return.

<SEQ=SEG.ACK><CTL=RST>

If the SEG.PRC is less than the TCB.PRC then continue.

Set RCV.NXT to SEG.SEQ+1, IRS is set to SEG.SEQ and any other control or text should be queued for processing later. ISS should be selected and a SYN segment sent of the form:

<SEQ=ISS><ACK=RCV.NXT><CTL=SYN,ACK>

SND.NXT is set to ISS+1 and SND.UNA to ISS. The connection state should be changed to SYN-RECEIVED. Note that any other incoming control or data (combined with SYN) will be processed in the SYN-RECEIVED state, but processing of SYN and ACK should not be repeated. If the listen was not fully specified (i.e., the foreign socket was not fully specified), then the unspecified fields should be filled in now.

fourth other text or control

Any other control or text-bearing segment (not containing SYN) must have an ACK and thus would be discarded by the ACK processing. An incoming RST segment could not be valid, since it could not have been sent in response to anything sent by this incarnation of the connection. So you are unlikely to get here, but if you do, drop the segment, and return.

If the state is SYN-SENT then

first check the ACK bit

If the ACK bit is set

If SEG.ACK =< ISS, or SEG.ACK > SND.NXT, send a reset (unless the RST bit is set, if so drop the segment and return)

<SEQ=SEG.ACK><CTL=RST>

and discard the segment. Return.

If SND.UNA =< SEG.ACK =< SND.NXT then the ACK is acceptable.

second check the RST bit

If the RST bit is set

If the ACK was acceptable then signal the user "error: connection reset", drop the segment, enter CLOSED state, delete TCB, and return. Otherwise (no ACK) drop the segment and return.

third check the security and precedence

If the security/compartment in the segment does not exactly match the security/compartment in the TCB, send a reset

If there is an ACK

<SEQ=SEG.ACK><CTL=RST>

Otherwise

<SEQ=0><ACK=SEG.SEQ+SEG.LEN><CTL=RST,ACK>

If there is an ACK

The precedence in the segment must match the precedence in the TCB, if not, send a reset

<SEQ=SEG.ACK><CTL=RST>

If there is no ACK

If the precedence in the segment is higher than the precedence in the TCB then if allowed by the user and the system raise the precedence in the TCB to that in the segment, if not allowed to raise the prec then send a reset.

<SEQ=0><ACK=SEG.SEQ+SEG.LEN><CTL=RST,ACK>

If the precedence in the segment is lower than the precedence in the TCB continue.

If a reset was sent, discard the segment and return.

fourth check the SYN bit

This step should be reached only if the ACK is ok, or there is no ACK, and it the segment did not contain a RST.

If the SYN bit is on and the security/compartment and precedence

RED HAT 1010
RED HAT V. CAS

are acceptable then, RCV.NXT is set to SEG.SEQ+1, IRS is set to SEG.SEQ. SND.UNA should be advanced to equal SEG.ACK (if there is an ACK), and any segments on the retransmission queue which are thereby acknowledged should be removed.

If SND.UNA > ISS (our SYN has been ACKed), change the connection state to ESTABLISHED, form an ACK segment

<SEQ=SND.NXT><ACK=RCV.NXT><CTL=ACK>

and send it. Data or controls which were queued for transmission may be included. If there are other controls or text in the segment then continue processing at the sixth step below where the URG bit is checked, otherwise return.

Otherwise enter SYN-RECEIVED, form a SYN,ACK segment

<SEQ=ISS><ACK=RCV.NXT><CTL=SYN,ACK>

and send it. If there are other controls or text in the segment, queue them for processing after the ESTABLISHED state has been reached, return.

fifth, if neither of the SYN or RST bits is set then drop the segment and return.

RED HAT 1010
RED HAT V. CAS

SEGMENT ARRIVES

Otherwise,

first check sequence number

SYN-RECEIVED STATE
ESTABLISHED STATE
FIN-WAIT-1 STATE
FIN-WAIT-2 STATE
CLOSE-WAIT STATE
CLOSING STATE
LAST-ACK STATE
TIME-WAIT STATE

Segments are processed in sequence. Initial tests on arrival are used to discard old duplicates, but further processing is done in SEG.SEQ order. If a segment's contents straddle the boundary between old and new, only the new parts should be processed.

There are four cases for the acceptability test for an incoming segment:

Segment Length	Receive Window	Test
0	0	SEG.SEQ = RCV.NXT
0	>0	RCV.NXT =< SEG.SEQ < RCV.NXT+RCV.WND
>0	0	not acceptable
>0	>0	RCV.NXT =< SEG.SEQ < RCV.NXT+RCV.WND or RCV.NXT =< SEG.SEQ+SEG.LEN-1 < RCV.NXT+RCV.WND

If the RCV.WND is zero, no segments will be acceptable, but special allowance should be made to accept valid ACKs, URGs and RSTs.

If an incoming segment is not acceptable, an acknowledgment should be sent in reply (unless the RST bit is set, if so drop the segment and return):

<SEQ=SND.NXT><ACK=RCV.NXT><CTL=ACK>

After sending the acknowledgment, drop the unacceptable segment and return.

RED HAT 1010
RED HAT V. CAS

In the following it is assumed that the segment is the idealized segment that begins at RCV.NXT and does not exceed the window. One could tailor actual segments to fit this assumption by trimming off any portions that lie outside the window (including SYN and FIN), and only processing further if the segment then begins at RCV.NXT. Segments with higher begining sequence numbers may be held for later processing.

second check the RST bit,

SYN-RECEIVED STATE

If the RST bit is set

If this connection was initiated with a passive OPEN (i.e., came from the LISTEN state), then return this connection to LISTEN state and return. The user need not be informed. If this connection was initiated with an active OPEN (i.e., came from SYN-SENT state) then the connection was refused, signal the user "connection refused". In either case, all segments on the retransmission queue should be removed. And in the active OPEN case, enter the CLOSED state and delete the TCB, and return.

ESTABLISHED
FIN-WAIT-1
FIN-WAIT-2
CLOSE-WAIT

If the RST bit is set then, any outstanding RECEIVEs and SEND should receive "reset" responses. All segment queues should be flushed. Users should also receive an unsolicited general "connection reset" signal. Enter the CLOSED state, delete the TCB, and return.

CLOSING STATE
LAST-ACK STATE
TIME-WAIT

If the RST bit is set then, enter the CLOSED state, delete the TCB, and return.

third check security and precedence

SYN-RECEIVED

If the security/compartment and precedence in the segment do not exactly match the security/compartment and precedence in the TCB then send a reset, and return.

ESTABLISHED STATE

If the security/compartment and precedence in the segment do not exactly match the security/compartment and precedence in the TCB then send a reset, any outstanding RECEIVEs and SEND should receive "reset" responses. All segment queues should be flushed. Users should also receive an unsolicited general "connection reset" signal. Enter the CLOSED state, delete the TCB, and return.

Note this check is placed following the sequence check to prevent a segment from an old connection between these ports with a different security or precedence from causing an abort of the current connection.

fourth, check the SYN bit,

SYN-RECEIVED
ESTABLISHED STATE
FIN-WAIT STATE-1
FIN-WAIT STATE-2
CLOSE-WAIT STATE
CLOSING STATE
LAST-ACK STATE
TIME-WAIT STATE

If the SYN is in the window it is an error, send a reset, any outstanding RECEIVEs and SEND should receive "reset" responses, all segment queues should be flushed, the user should also receive an unsolicited general "connection reset" signal, enter the CLOSED state, delete the TCB, and return.

If the SYN is not in the window this step would not be reached and an ack would have been sent in the first step (sequence number check).

RED HAT 1010
RED HAT V. CAS

fifth check the ACK field,

if the ACK bit is off drop the segment and return

if the ACK bit is on

SYN-RECEIVED STATE

If SND.UNA =< SEG.ACK =< SND.NXT then enter ESTABLISHED state and continue processing.

If the segment acknowledgment is not acceptable, form a reset segment,

<SEQ=SEG.ACK><CTL=RST>

and send it.

ESTABLISHED STATE

If SND.UNA < SEG.ACK =< SND.NXT then, set SND.UNA <- SEG.ACK. Any segments on the retransmission queue which are thereby entirely acknowledged are removed. Users should receive positive acknowledgments for buffers which have been SENT and fully acknowledged (i.e., SEND buffer should be returned with "ok" response). If the ACK is a duplicate (SEG.ACK < SND.UNA), it can be ignored. If the ACK acks something not yet sent (SEG.ACK > SND.NXT) then send an ACK, drop the segment, and return.

If SND.UNA < SEG.ACK =< SND.NXT, the send window should be updated. If (SND.WL1 < SEG.SEQ or (SND.WL1 = SEG.SEQ and SND.WL2 =< SEG.ACK)), set SND.WND <- SEG.WND, set SND.WL1 <- SEG.SEQ, and set SND.WL2 <- SEG.ACK.

Note that SND.WND is an offset from SND.UNA, that SND.WL1 records the sequence number of the last segment used to update SND.WND, and that SND.WL2 records the acknowledgment number of the last segment used to update SND.WND. The check here prevents using old segments to update the window.

RED HAT 1010
RED HAT V. CAS

SEGMENT ARRIVES

FIN-WAIT-1 STATE

In addition to the processing for the ESTABLISHED state, if our FIN is now acknowledged then enter FIN-WAIT-2 and continue processing in that state.

FIN-WAIT-2 STATE

In addition to the processing for the ESTABLISHED state, if the retransmission queue is empty, the user's CLOSE can be acknowledged ("ok") but do not delete the TCB.

CLOSE-WAIT STATE

Do the same processing as for the ESTABLISHED state.

CLOSING STATE

In addition to the processing for the ESTABLISHED state, if the ACK acknowledges our FIN then enter the TIME-WAIT state, otherwise ignore the segment.

LAST-ACK STATE

The only thing that can arrive in this state is an acknowledgment of our FIN. If our FIN is now acknowledged, delete the TCB, enter the CLOSED state, and return.

TIME-WAIT STATE

The only thing that can arrive in this state is a retransmission of the remote FIN. Acknowledge it, and restart the 2 MSL timeout.

sixth, check the URG bit,

ESTABLISHED STATE
FIN-WAIT-1 STATE
FIN-WAIT-2 STATE

If the URG bit is set, RCV.UP <- max(RCV.UP,SEG.UP), and signal the user that the remote side has urgent data if the urgent pointer (RCV.UP) is in advance of the data consumed. If the user has already been signaled (or is still in the "urgent mode") for this continuous sequence of urgent data, do not signal the user again.

RED HAT 1010
RED HAT V. CAS

CLOSE-WAIT STATE
CLOSING STATE
LAST-ACK STATE
TIME-WAIT

This should not occur, since a FIN has been received from the remote side. Ignore the URG.

seventh, process the segment text,

ESTABLISHED STATE
FIN-WAIT-1 STATE
FIN-WAIT-2 STATE

Once in the ESTABLISHED state, it is possible to deliver segment text to user RECEIVE buffers. Text from segments can be moved into buffers until either the buffer is full or the segment is empty. If the segment empties and carries an PUSH flag, then the user is informed, when the buffer is returned, that a PUSH has been received.

When the TCP takes responsibility for delivering the data to the user it must also acknowledge the receipt of the data.

Once the TCP takes responsibility for the data it advances RCV.NXT over the data accepted, and adjusts RCV.WND as apporopriate to the current buffer availability. The total of RCV.NXT and RCV.WND should not be reduced.

Please note the window management suggestions in section 3.7.

Send an acknowledgment of the form:

<SEQ=SND.NXT><ACK=RCV.NXT><CTL=ACK>

This acknowledgment should be piggybacked on a segment being transmitted if possible without incurring undue delay.

CLOSE-WAIT STATE
CLOSING STATE
LAST-ACK STATE
TIME-WAIT STATE

This should not occur, since a FIN has been received from the remote side. Ignore the segment text.

eighth, check the FIN bit,

Do not process the FIN if the state is CLOSED, LISTEN or SYN-SENT since the SEG.SEQ cannot be validated; drop the segment and return.

If the FIN bit is set, signal the user "connection closing" and return any pending RECEIVEs with same message, advance RCV.NXT over the FIN, and send an acknowledgment for the FIN. Note that FIN implies PUSH for any segment text not yet delivered to the user.

SYN-RECEIVED STATE
ESTABLISHED STATE

Enter the CLOSE-WAIT state.

FIN-WAIT-1 STATE

If our FIN has been ACKed (perhaps in this segment), then enter TIME-WAIT, start the time-wait timer, turn off the other timers; otherwise enter the CLOSING state.

FIN-WAIT-2 STATE

Enter the TIME-WAIT state. Start the time-wait timer, turn off the other timers.

CLOSE-WAIT STATE

Remain in the CLOSE-WAIT state.

CLOSING STATE

Remain in the CLOSING state.

LAST-ACK STATE

Remain in the LAST-ACK state.

RED HAT 1010
RED HAT V. CAS

TIME-WAIT STATE

Remain in the TIME-WAIT state. Restart the 2 MSL time-wait timeout.

and return.

USER TIMEOUT

For any state if the user timeout expires, flush all queues, signal the user "error: connection aborted due to user timeout" in general and for any outstanding calls, delete the TCB, enter the CLOSED state and return.

RETRANSMISSION TIMEOUT

For any state if the retransmission timeout expires on a segment in the retransmission queue, send the segment at the front of the retransmission queue again, reinitialize the retransmission timer, and return.

TIME-WAIT TIMEOUT

If the time-wait timeout expires on a connection delete the TCB, enter the CLOSED state and return.

RED HAT 1010
RED HAT V. CAS

RED HAT 1010
RED HAT V. CAS

GLOSSARY

1822
: BBN Report 1822, "The Specification of the Interconnection of a Host and an IMP". The specification of interface between a host and the ARPANET.

ACK
: A control bit (acknowledge) occupying no sequence space, which indicates that the acknowledgment field of this segment specifies the next sequence number the sender of this segment is expecting to receive, hence acknowledging receipt of all previous sequence numbers.

ARPANET message
: The unit of transmission between a host and an IMP in the ARPANET. The maximum size is about 1012 octets (8096 bits).

ARPANET packet
: A unit of transmission used internally in the ARPANET between IMPs. The maximum size is about 126 octets (1008 bits).

connection
: A logical communication path identified by a pair of sockets.

datagram
: A message sent in a packet switched computer communications network.

Destination Address
: The destination address, usually the network and host identifiers.

FIN
: A control bit (finis) occupying one sequence number, which indicates that the sender will send no more data or control occupying sequence space.

fragment
: A portion of a logical unit of data, in particular an internet fragment is a portion of an internet datagram.

FTP
: A file transfer protocol.

RED HAT 1010
RED HAT V. CAS

header
: Control information at the beginning of a message, segment, fragment, packet or block of data.

host
: A computer. In particular a source or destination of messages from the point of view of the communication network.

Identification
: An Internet Protocol field. This identifying value assigned by the sender aids in assembling the fragments of a datagram.

IMP
: The Interface Message Processor, the packet switch of the ARPANET.

internet address
: A source or destination address specific to the host level.

internet datagram
: The unit of data exchanged between an internet module and the higher level protocol together with the internet header.

internet fragment
: A portion of the data of an internet datagram with an internet header.

IP
: Internet Protocol.

IRS
: The Initial Receive Sequence number. The first sequence number used by the sender on a connection.

ISN
: The Initial Sequence Number. The first sequence number used on a connection, (either ISS or IRS). Selected on a clock based procedure.

ISS
: The Initial Send Sequence number. The first sequence number used by the sender on a connection.

leader
: Control information at the beginning of a message or block of data. In particular, in the ARPANET, the control information on an ARPANET message at the host-IMP interface.

RED HAT 1010
RED HAT V. CAS

left sequence
: This is the next sequence number to be acknowledged by the data receiving TCP (or the lowest currently unacknowledged sequence number) and is sometimes referred to as the left edge of the send window.

local packet
: The unit of transmission within a local network.

module
: An implementation, usually in software, of a protocol or other procedure.

MSL
: Maximum Segment Lifetime, the time a TCP segment can exist in the internetwork system. Arbitrarily defined to be 2 minutes.

octet
: An eight bit byte.

Options
: An Option field may contain several options, and each option may be several octets in length. The options are used primarily in testing situations; for example, to carry timestamps. Both the Internet Protocol and TCP provide for options fields.

packet
: A package of data with a header which may or may not be logically complete. More often a physical packaging than a logical packaging of data.

port
: The portion of a socket that specifies which logical input or output channel of a process is associated with the data.

process
: A program in execution. A source or destination of data from the point of view of the TCP or other host-to-host protocol.

PUSH
: A control bit occupying no sequence space, indicating that this segment contains data that must be pushed through to the receiving user.

RCV.NXT
: receive next sequence number

RED HAT 1010
RED HAT V. CAS

RCV.UP
: receive urgent pointer

RCV.WND
: receive window

receive next sequence number
: This is the next sequence number the local TCP is expecting to receive.

receive window
: This represents the sequence numbers the local (receiving) TCP is willing to receive. Thus, the local TCP considers that segments overlapping the range RCV.NXT to RCV.NXT + RCV.WND - 1 carry acceptable data or control. Segments containing sequence numbers entirely outside of this range are considered duplicates and discarded.

RST
: A control bit (reset), occupying no sequence space, indicating that the receiver should delete the connection without further interaction. The receiver can determine, based on the sequence number and acknowledgment fields of the incoming segment, whether it should honor the reset command or ignore it. In no case does receipt of a segment containing RST give rise to a RST in response.

RTP
: Real Time Protocol: A host-to-host protocol for communication of time critical information.

SEG.ACK
: segment acknowledgment

SEG.LEN
: segment length

SEG.PRC
: segment precedence value

SEG.SEQ
: segment sequence

SEG.UP
: segment urgent pointer field

SEG.WND
segment window field

segment
A logical unit of data, in particular a TCP segment is the unit of data transfered between a pair of TCP modules.

segment acknowledgment
The sequence number in the acknowledgment field of the arriving segment.

segment length
The amount of sequence number space occupied by a segment, including any controls which occupy sequence space.

segment sequence
The number in the sequence field of the arriving segment.

send sequence
This is the next sequence number the local (sending) TCP will use on the connection. It is initially selected from an initial sequence number curve (ISN) and is incremented for each octet of data or sequenced control transmitted.

send window
This represents the sequence numbers which the remote (receiving) TCP is willing to receive. It is the value of the window field specified in segments from the remote (data receiving) TCP. The range of new sequence numbers which may be emitted by a TCP lies between SND.NXT and SND.UNA + SND.WND - 1. (Retransmissions of sequence numbers between SND.UNA and SND.NXT are expected, of course.)

SND.NXT
send sequence

SND.UNA
left sequence

SND.UP
send urgent pointer

SND.WL1
segment sequence number at last window update

SND.WL2
segment acknowledgment number at last window update

RED HAT 1010
RED HAT V. CAS

SND.WND
: send window

socket
: An address which specifically includes a port identifier, that is, the concatenation of an Internet Address with a TCP port.

Source Address
: The source address, usually the network and host identifiers.

SYN
: A control bit in the incoming segment, occupying one sequence number, used at the initiation of a connection, to indicate where the sequence numbering will start.

TCB
: Transmission control block, the data structure that records the state of a connection.

TCB.PRC
: The precedence of the connection.

TCP
: Transmission Control Protocol: A host-to-host protocol for reliable communication in internetwork environments.

TOS
: Type of Service, an Internet Protocol field.

Type of Service
: An Internet Protocol field which indicates the type of service for this internet fragment.

URG
: A control bit (urgent), occupying no sequence space, used to indicate that the receiving user should be notified to do urgent processing as long as there is data to be consumed with sequence numbers less than the value indicated in the urgent pointer.

urgent pointer
: A control field meaningful only when the URG bit is on. This field communicates the value of the urgent pointer which indicates the data octet associated with the sending user's urgent call.

RED HAT 1010
RED HAT V. CAS

REFERENCES

[1] Cerf, V., and R. Kahn, "A Protocol for Packet Network Intercommunication", IEEE Transactions on Communications, Vol. COM-22, No. 5, pp 637-648, May 1974.

[2] Postel, J. (ed.), "Internet Protocol - DARPA Internet Program Protocol Specification", RFC 791, USC/Information Sciences Institute, September 1981.

[3] Dalal, Y. and C. Sunshine, "Connection Management in Transport Protocols", Computer Networks, Vol. 2, No. 6, pp. 454-473, December 1978.

[4] Postel, J., "Assigned Numbers", RFC 790, USC/Information Sciences Institute, September 1981.

RED HAT 1010
RED HAT V. CAS

Network Working Group
Request for Comments: 1323
Obsoletes: RFC 1072, RFC 1185

V. Jacobson
LBL
R. Braden
ISI
D. Borman
Cray Research

May 1992

TCP Extensions for High Performance

Status of This Memo

This RFC specifies an IAB standards track protocol for the Internet community, and requests discussion and suggestions for improvements. Please refer to the current edition of the "IAB Official Protocol Standards" for the standardization state and status of this protocol. Distribution of this memo is unlimited.

Abstract

This memo presents a set of TCP extensions to improve performance over large bandwidth*delay product paths and to provide reliable operation over very high-speed paths. It defines new TCP options for scaled windows and timestamps, which are designed to provide compatible interworking with TCP's that do not implement the extensions. The timestamps are used for two distinct mechanisms: RTTM (Round Trip Time Measurement) and PAWS (Protect Against Wrapped Sequences). Selective acknowledgments are not included in this memo.

This memo combines and supersedes RFC-1072 and RFC-1185, adding additional clarification and more detailed specification. Appendix C summarizes the changes from the earlier RFCs.

TABLE OF CONTENTS

1. Introduction 2
2. TCP Window Scale Option 8
3. RTTM -- Round-Trip Time Measurement 11
4. PAWS -- Protect Against Wrapped Sequence Numbers 17
5. Conclusions and Acknowledgments 25
6. References 25
APPENDIX A: Implementation Suggestions 27
APPENDIX B: Duplicates from Earlier Connection Incarnations 27
APPENDIX C: Changes from RFC-1072, RFC-1185 30
APPENDIX D: Summary of Notation 31
APPENDIX E: Event Processing 32
Security Considerations 37

RED HAT 1010
RED HAT V. CAS

Authors' Addresses 37

1. INTRODUCTION

The TCP protocol [Postel81] was designed to operate reliably over almost any transmission medium regardless of transmission rate, delay, corruption, duplication, or reordering of segments. Production TCP implementations currently adapt to transfer rates in the range of 100 bps to 10**7 bps and round-trip delays in the range 1 ms to 100 seconds. Recent work on TCP performance has shown that TCP can work well over a variety of Internet paths, ranging from 800 Mbit/sec I/O channels to 300 bit/sec dial-up modems [Jacobson88a].

The introduction of fiber optics is resulting in ever-higher transmission speeds, and the fastest paths are moving out of the domain for which TCP was originally engineered. This memo defines a set of modest extensions to TCP to extend the domain of its application to match this increasing network capability. It is based upon and obsoletes RFC-1072 [Jacobson88b] and RFC-1185 [Jacobson90b].

There is no one-line answer to the question: "How fast can TCP go?". There are two separate kinds of issues, performance and reliability, and each depends upon different parameters. We discuss each in turn.

1.1 TCP Performance

TCP performance depends not upon the transfer rate itself, but rather upon the product of the transfer rate and the round-trip delay. This "bandwidth*delay product" measures the amount of data that would "fill the pipe"; it is the buffer space required at sender and receiver to obtain maximum throughput on the TCP connection over the path, i.e., the amount of unacknowledged data that TCP must handle in order to keep the pipeline full. TCP performance problems arise when the bandwidth*delay product is large. We refer to an Internet path operating in this region as a "long, fat pipe", and a network containing this path as an "LFN" (pronounced "elephan(t)").

High-capacity packet satellite channels (e.g., DARPA's Wideband Net) are LFN's. For example, a DS1-speed satellite channel has a bandwidth*delay product of 10**6 bits or more; this corresponds to 100 outstanding TCP segments of 1200 bytes each. Terrestrial fiber-optical paths will also fall into the LFN class; for example, a cross-country delay of 30 ms at a DS3 bandwidth (45Mbps) also exceeds 10**6 bits.

There are three fundamental performance problems with the current TCP over LFN paths:

RED HAT 1010
RED HAT V. CAS

(1) Window Size Limit

The TCP header uses a 16 bit field to report the receive window size to the sender. Therefore, the largest window that can be used is 2**16 = 65K bytes.

To circumvent this problem, Section 2 of this memo defines a new TCP option, "Window Scale", to allow windows larger than 2**16. This option defines an implicit scale factor, which is used to multiply the window size value found in a TCP header to obtain the true window size.

(2) Recovery from Losses

Packet losses in an LFN can have a catastrophic effect on throughput. Until recently, properly-operating TCP implementations would cause the data pipeline to drain with every packet loss, and require a slow-start action to recover. Recently, the Fast Retransmit and Fast Recovery algorithms [Jacobson90c] have been introduced. Their combined effect is to recover from one packet loss per window, without draining the pipeline. However, more than one packet loss per window typically results in a retransmission timeout and the resulting pipeline drain and slow start.

Expanding the window size to match the capacity of an LFN results in a corresponding increase of the probability of more than one packet per window being dropped. This could have a devastating effect upon the throughput of TCP over an LFN. In addition, if a congestion control mechanism based upon some form of random dropping were introduced into gateways, randomly spaced packet drops would become common, possible increasing the probability of dropping more than one packet per window.

To generalize the Fast Retransmit/Fast Recovery mechanism to handle multiple packets dropped per window, selective acknowledgments are required. Unlike the normal cumulative acknowledgments of TCP, selective acknowledgments give the sender a complete picture of which segments are queued at the receiver and which have not yet arrived. Some evidence in favor of selective acknowledgments has been published [NBS85], and selective acknowledgments have been included in a number of experimental Internet protocols -- VMTP [Cheriton88], NETBLT [Clark87], and RDP [Velten84], and proposed for OSI TP4 [NBS85]. However, in the non-LFN regime, selective acknowledgments reduce the number of

RED HAT 1010
RED HAT V. CAS

packets retransmitted but do not otherwise improve performance, making their complexity of questionable value. However, selective acknowledgments are expected to become much more important in the LFN regime.

RFC-1072 defined a new TCP "SACK" option to send a selective acknowledgment. However, there are important technical issues to be worked out concerning both the format and semantics of the SACK option. Therefore, SACK has been omitted from this package of extensions. It is hoped that SACK can "catch up" during the standardization process.

(3) Round-Trip Measurement

TCP implements reliable data delivery by retransmitting segments that are not acknowledged within some retransmission timeout (RTO) interval. Accurate dynamic determination of an appropriate RTO is essential to TCP performance. RTO is determined by estimating the mean and variance of the measured round-trip time (RTT), i.e., the time interval between sending a segment and receiving an acknowledgment for it [Jacobson88a].

Section 4 introduces a new TCP option, "Timestamps", and then defines a mechanism using this option that allows nearly every segment, including retransmissions, to be timed at negligible computational cost. We use the mnemonic RTTM (Round Trip Time Measurement) for this mechanism, to distinguish it from other uses of the Timestamps option.

1.2 TCP Reliability

Now we turn from performance to reliability. High transfer rate enters TCP performance through the bandwidth*delay product. However, high transfer rate alone can threaten TCP reliability by violating the assumptions behind the TCP mechanism for duplicate detection and sequencing.

An especially serious kind of error may result from an accidental reuse of TCP sequence numbers in data segments. Suppose that an "old duplicate segment", e.g., a duplicate data segment that was delayed in Internet queues, is delivered to the receiver at the wrong moment, so that its sequence numbers falls somewhere within the current window. There would be no checksum failure to warn of the error, and the result could be an undetected corruption of the data. Reception of an old duplicate ACK segment at the transmitter could be only slightly less serious: it is likely to

RED HAT 1010
RED HAT V. CAS

lock up the connection so that no further progress can be made, forcing an RST on the connection.

TCP reliability depends upon the existence of a bound on the lifetime of a segment: the "Maximum Segment Lifetime" or MSL. An MSL is generally required by any reliable transport protocol, since every sequence number field must be finite, and therefore any sequence number may eventually be reused. In the Internet protocol suite, the MSL bound is enforced by an IP-layer mechanism, the "Time-to-Live" or TTL field.

Duplication of sequence numbers might happen in either of two ways:

(1) Sequence number wrap-around on the current connection

A TCP sequence number contains 32 bits. At a high enough transfer rate, the 32-bit sequence space may be "wrapped" (cycled) within the time that a segment is delayed in queues.

(2) Earlier incarnation of the connection

Suppose that a connection terminates, either by a proper close sequence or due to a host crash, and the same connection (i.e., using the same pair of sockets) is immediately reopened. A delayed segment from the terminated connection could fall within the current window for the new incarnation and be accepted as valid.

Duplicates from earlier incarnations, Case (2), are avoided by enforcing the current fixed MSL of the TCP spec, as explained in Section 5.3 and Appendix B. However, case (1), avoiding the reuse of sequence numbers within the same connection, requires an MSL bound that depends upon the transfer rate, and at high enough rates, a new mechanism is required.

More specifically, if the maximum effective bandwidth at which TCP is able to transmit over a particular path is B bytes per second, then the following constraint must be satisfied for error-free operation:

```
2**31 / B  > MSL (secs)                    [1]
```

The following table shows the value for Twrap = 2**31/B in seconds, for some important values of the bandwidth B:

RED HAT 1010
RED HAT V. CAS

Network	B*8 bits/sec	B bytes/sec	Twrap secs
ARPANET	56kbps	7KBps	3*10**5 (~3.6 days)
DS1	1.5Mbps	190KBps	10**4 (~3 hours)
Ethernet	10Mbps	1.25MBps	1700 (~30 mins)
DS3	45Mbps	5.6MBps	380
FDDI	100Mbps	12.5MBps	170
Gigabit	1Gbps	125MBps	17

It is clear that wrap-around of the sequence space is not a problem for 56kbps packet switching or even 10Mbps Ethernets. On the other hand, at DS3 and FDDI speeds, Twrap is comparable to the 2 minute MSL assumed by the TCP specification [Postel81]. Moving towards gigabit speeds, Twrap becomes too small for reliable enforcement by the Internet TTL mechanism.

The 16-bit window field of TCP limits the effective bandwidth B to 2**16/RTT, where RTT is the round-trip time in seconds [McKenzie89]. If the RTT is large enough, this limits B to a value that meets the constraint [1] for a large MSL value. For example, consider a transcontinental backbone with an RTT of 60ms (set by the laws of physics). With the bandwidth*delay product limited to 64KB by the TCP window size, B is then limited to 1.1MBps, no matter how high the theoretical transfer rate of the path. This corresponds to cycling the sequence number space in Twrap= 2000 secs, which is safe in today's Internet.

It is important to understand that the culprit is not the larger window but rather the high bandwidth. For example, consider a (very large) FDDI LAN with a diameter of 10km. Using the speed of light, we can compute the RTT across the ring as (2*10**4)/(3*10**8) = 67 microseconds, and the delay*bandwidth product is then 833 bytes. A TCP connection across this LAN using a window of only 833 bytes will run at the full 100mbps and can wrap the sequence space in about 3 minutes, very close to the MSL of TCP. Thus, high speed alone can cause a reliability problem with sequence number wrap-around, even without extended windows.

Watson's Delta-T protocol [Watson81] includes network-layer mechanisms for precise enforcement of an MSL. In contrast, the IP

RED HAT 1010
RED HAT V. CAS

mechanism for MSL enforcement is loosely defined and even more loosely implemented in the Internet. Therefore, it is unwise to depend upon active enforcement of MSL for TCP connections, and it is unrealistic to imagine setting MSL's smaller than the current values (e.g., 120 seconds specified for TCP).

A possible fix for the problem of cycling the sequence space would be to increase the size of the TCP sequence number field. For example, the sequence number field (and also the acknowledgment field) could be expanded to 64 bits. This could be done either by changing the TCP header or by means of an additional option.

Section 5 presents a different mechanism, which we call PAWS (Protect Against Wrapped Sequence numbers), to extend TCP reliability to transfer rates well beyond the foreseeable upper limit of network bandwidths. PAWS uses the TCP Timestamps option defined in Section 4 to protect against old duplicates from the same connection.

1.3 Using TCP options

The extensions defined in this memo all use new TCP options. We must address two possible issues concerning the use of TCP options: (1) compatibility and (2) overhead.

We must pay careful attention to compatibility, i.e., to interoperation with existing implementations. The only TCP option defined previously, MSS, may appear only on a SYN segment. Every implementation should (and we expect that most will) ignore unknown options on SYN segments. However, some buggy TCP implementation might be crashed by the first appearance of an option on a non-SYN segment. Therefore, for each of the extensions defined below, TCP options will be sent on non-SYN segments only when an exchange of options on the SYN segments has indicated that both sides understand the extension. Furthermore, an extension option will be sent in a <SYN,ACK> segment only if the corresponding option was received in the initial <SYN> segment.

A question may be raised about the bandwidth and processing overhead for TCP options. Those options that occur on SYN segments are not likely to cause a performance concern. Opening a TCP connection requires execution of significant special-case code, and the processing of options is unlikely to increase that cost significantly.

On the other hand, a Timestamps option may appear in any data or ACK segment, adding 12 bytes to the 20-byte TCP header. We

RED HAT 1010
RED HAT V. CAS

believe that the bandwidth saved by reducing unnecessary retransmissions will more than pay for the extra header bandwidth.

There is also an issue about the processing overhead for parsing the variable byte-aligned format of options, particularly with a RISC-architecture CPU. To meet this concern, Appendix A contains a recommended layout of the options in TCP headers to achieve reasonable data field alignment. In the spirit of Header Prediction, a TCP can quickly test for this layout and if it is verified then use a fast path. Hosts that use this canonical layout will effectively use the options as a set of fixed-format fields appended to the TCP header. However, to retain the philosophical and protocol framework of TCP options, a TCP must be prepared to parse an arbitrary options field, albeit with less efficiency.

Finally, we observe that most of the mechanisms defined in this memo are important for LFN's and/or very high-speed networks. For low-speed networks, it might be a performance optimization to NOT use these mechanisms. A TCP vendor concerned about optimal performance over low-speed paths might consider turning these extensions off for low-speed paths, or allow a user or installation manager to disable them.

2. TCP WINDOW SCALE OPTION

2.1 Introduction

The window scale extension expands the definition of the TCP window to 32 bits and then uses a scale factor to carry this 32-bit value in the 16-bit Window field of the TCP header (SEG.WND in RFC-793). The scale factor is carried in a new TCP option, Window Scale. This option is sent only in a SYN segment (a segment with the SYN bit on), hence the window scale is fixed in each direction when a connection is opened. (Another design choice would be to specify the window scale in every TCP segment. It would be incorrect to send a window scale option only when the scale factor changed, since a TCP option in an acknowledgement segment will not be delivered reliably (unless the ACK happens to be piggy-backed on data in the other direction). Fixing the scale when the connection is opened has the advantage of lower overhead but the disadvantage that the scale factor cannot be changed during the connection.)

The maximum receive window, and therefore the scale factor, is determined by the maximum receive buffer space. In a typical modern implementation, this maximum buffer space is set by default

RED HAT 1010
303
RED HAT V. CAS

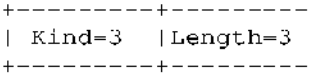
but can be overridden by a user program before a TCP connection is opened. This determines the scale factor, and therefore no new user interface is needed for window scaling.

2.2 Window Scale Option

The three-byte Window Scale option may be sent in a SYN segment by a TCP. It has two purposes: (1) indicate that the TCP is prepared to do both send and receive window scaling, and (2) communicate a scale factor to be applied to its receive window. Thus, a TCP that is prepared to scale windows should send the option, even if its own scale factor is 1. The scale factor is limited to a power of two and encoded logarithmically, so it may be implemented by binary shift operations.

TCP Window Scale Option (WSopt):

Kind: 3 Length: 3 bytes



This option is an offer, not a promise; both sides must send Window Scale options in their SYN segments to enable window scaling in either direction. If window scaling is enabled, then the TCP that sent this option will right-shift its true receive-window values by 'shift.cnt' bits for transmission in SEG.WND. The value 'shift.cnt' may be zero (offering to scale, while applying a scale factor of 1 to the receive window).

This option may be sent in an initial <SYN> segment (i.e., a segment with the SYN bit on and the ACK bit off). It may also be sent in a <SYN,ACK> segment, but only if a Window Scale option was received in the initial <SYN> segment. A Window Scale option in a segment without a SYN bit should be ignored.

The Window field in a SYN (i.e., a <SYN> or <SYN,ACK>) segment itself is never scaled.

2.3 Using the Window Scale Option

A model implementation of window scaling is as follows, using the notation of RFC-793 [Postel81]:

* All windows are treated as 32-bit quantities for storage in

RED HAT 1010
RED HAT V. CAS

the connection control block and for local calculations. This includes the send-window (SND.WND) and the receive-window (RCV.WND) values, as well as the congestion window.

* The connection state is augmented by two window shift counts, Snd.Wind.Scale and Rcv.Wind.Scale, to be applied to the incoming and outgoing window fields, respectively.

* If a TCP receives a <SYN> segment containing a Window Scale option, it sends its own Window Scale option in the <SYN,ACK> segment.

* The Window Scale option is sent with shift.cnt = R, where R is the value that the TCP would like to use for its receive window.

* Upon receiving a SYN segment with a Window Scale option containing shift.cnt = S, a TCP sets Snd.Wind.Scale to S and sets Rcv.Wind.Scale to R; otherwise, it sets both Snd.Wind.Scale and Rcv.Wind.Scale to zero.

* The window field (SEG.WND) in the header of every incoming segment, with the exception of SYN segments, is left-shifted by Snd.Wind.Scale bits before updating SND.WND:

    SND.WND = SEG.WND << Snd.Wind.Scale

    (assuming the other conditions of RFC793 are met, and using the "C" notation "<<" for left-shift).

* The window field (SEG.WND) of every outgoing segment, with the exception of SYN segments, is right-shifted by Rcv.Wind.Scale bits:

    SEG.WND = RCV.WND >> Rcv.Wind.Scale.

TCP determines if a data segment is "old" or "new" by testing whether its sequence number is within 2**31 bytes of the left edge of the window, and if it is not, discarding the data as "old". To insure that new data is never mistakenly considered old and vice-versa, the left edge of the sender's window has to be at most 2**31 away from the right edge of the receiver's window. Similarly with the sender's right edge and receiver's left edge. Since the right and left edges of either the sender's or receiver's window differ by the window size, and since the sender and receiver windows can be out of phase by at most the window size, the above constraints imply that 2 * the max window size

RED HAT 1010
RED HAT V. CAS

must be less than 2**31, or

max window < 2**30

Since the max window is 2**S (where S is the scaling shift count) times at most 2**16 - 1 (the maximum unscaled window), the maximum window is guaranteed to be < 2*30 if S <= 14. Thus, the shift count must be limited to 14 (which allows windows of 2**30 = 1 Gbyte). If a Window Scale option is received with a shift.cnt value exceeding 14, the TCP should log the error but use 14 instead of the specified value.

The scale factor applies only to the Window field as transmitted in the TCP header; each TCP using extended windows will maintain the window values locally as 32-bit numbers. For example, the "congestion window" computed by Slow Start and Congestion Avoidance is not affected by the scale factor, so window scaling will not introduce quantization into the congestion window.

3. RTTM: ROUND-TRIP TIME MEASUREMENT

3.1 Introduction

Accurate and current RTT estimates are necessary to adapt to changing traffic conditions and to avoid an instability known as "congestion collapse" [Nagle84] in a busy network. However, accurate measurement of RTT may be difficult both in theory and in implementation.

Many TCP implementations base their RTT measurements upon a sample of only one packet per window. While this yields an adequate approximation to the RTT for small windows, it results in an unacceptably poor RTT estimate for an LFN. If we look at RTT estimation as a signal processing problem (which it is), a data signal at some frequency, the packet rate, is being sampled at a lower frequency, the window rate. This lower sampling frequency violates Nyquist's criteria and may therefore introduce "aliasing" artifacts into the estimated RTT [Hamming77].

A good RTT estimator with a conservative retransmission timeout calculation can tolerate aliasing when the sampling frequency is "close" to the data frequency. For example, with a window of 8 packets, the sample rate is 1/8 the data frequency -- less than an order of magnitude different. However, when the window is tens or hundreds of packets, the RTT estimator may be seriously in error, resulting in spurious retransmissions.

If there are dropped packets, the problem becomes worse. Zhang

[Zhang86], Jain [Jain86] and Karn [Karn87] have shown that it is not possible to accumulate reliable RTT estimates if retransmitted segments are included in the estimate. Since a full window of data will have been transmitted prior to a retransmission, all of the segments in that window will have to be ACKed before the next RTT sample can be taken. This means at least an additional window's worth of time between RTT measurements and, as the error rate approaches one per window of data (e.g., 10**-6 errors per bit for the Wideband satellite network), it becomes effectively impossible to obtain a valid RTT measurement.

A solution to these problems, which actually simplifies the sender substantially, is as follows: using TCP options, the sender places a timestamp in each data segment, and the receiver reflects these timestamps back in ACK segments. Then a single subtract gives the sender an accurate RTT measurement for every ACK segment (which will correspond to every other data segment, with a sensible receiver). We call this the RTTM (Round-Trip Time Measurement) mechanism.

It is vitally important to use the RTTM mechanism with big windows; otherwise, the door is opened to some dangerous instabilities due to aliasing. Furthermore, the option is probably useful for all TCP's, since it simplifies the sender.

3.2 TCP Timestamps Option

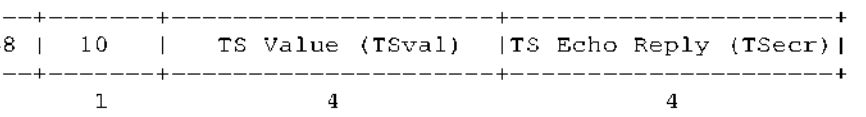
TCP is a symmetric protocol, allowing data to be sent at any time in either direction, and therefore timestamp echoing may occur in either direction. For simplicity and symmetry, we specify that timestamps always be sent and echoed in both directions. For efficiency, we combine the timestamp and timestamp reply fields into a single TCP Timestamps Option.

RED HAT 1010
RED HAT V. CAS

TCP Timestamps Option (TSopt):

Kind: 8

Length: 10 bytes



The Timestamps option carries two four-byte timestamp fields. The Timestamp Value field (TSval) contains the current value of the timestamp clock of the TCP sending the option.

The Timestamp Echo Reply field (TSecr) is only valid if the ACK bit is set in the TCP header; if it is valid, it echos a timestamp value that was sent by the remote TCP in the TSval field of a Timestamps option. When TSecr is not valid, its value must be zero. The TSecr value will generally be from the most recent Timestamp option that was received; however, there are exceptions that are explained below.

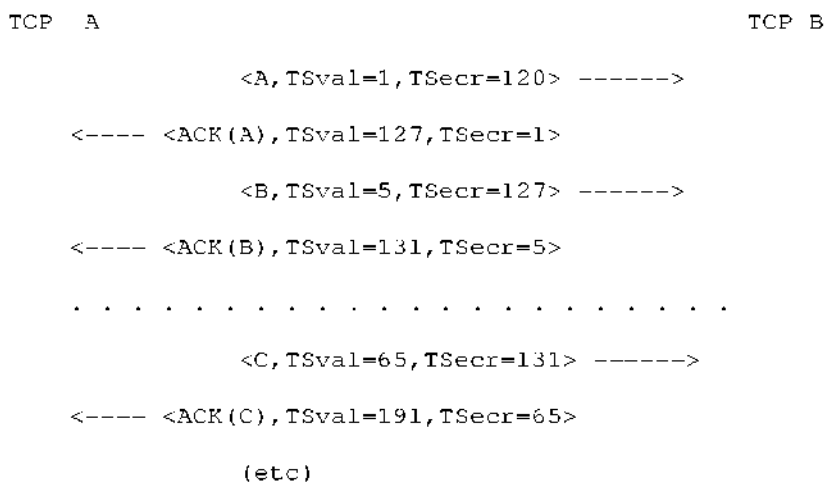
A TCP may send the Timestamps option (TSopt) in an initial <SYN> segment (i.e., segment containing a SYN bit and no ACK bit), and may send a TSopt in other segments only if it received a TSopt in the initial <SYN> segment for the connection.

3.3 The RTTM Mechanism

The timestamp value to be sent in TSval is to be obtained from a (virtual) clock that we call the "timestamp clock". Its values must be at least approximately proportional to real time, in order to measure actual RTT.

The following example illustrates a one-way data flow with segments arriving in sequence without loss. Here A, B, C... represent data blocks occupying successive blocks of sequence numbers, and ACK(A),... represent the corresponding cumulative acknowledgments. The two timestamp fields of the Timestamps option are shown symbolically as <TSval= x,TSecr=y>. Each TSecr field contains the value most recently received in a TSval field.

RED HAT 1010
RED HAT V. CAS



The dotted line marks a pause (60 time units long) in which A had nothing to send. Note that this pause inflates the RTT which B could infer from receiving TSecr=131 in data segment C. Thus, in one-way data flows, RTTM in the reverse direction measures a value that is inflated by gaps in sending data. However, the following rule prevents a resulting inflation of the measured RTT:

> A TSecr value received in a segment is used to update the averaged RTT measurement only if the segment acknowledges some new data, i.e., only if it advances the left edge of the send window.

Since TCP B is not sending data, the data segment C does not acknowledge any new data when it arrives at B. Thus, the inflated RTTM measurement is not used to update B's RTTM measurement.

3.4 Which Timestamp to Echo

If more than one Timestamps option is received before a reply segment is sent, the TCP must choose only one of the TSvals to echo, ignoring the others. To minimize the state kept in the receiver (i.e., the number of unprocessed TSvals), the receiver should be required to retain at most one timestamp in the connection control block.

There are three situations to consider:

(A) Delayed ACKs.

Many TCP's acknowledge only every Kth segment out of a group of segments arriving within a short time interval; this policy is known generally as "delayed ACKs". The data-sender TCP must measure the effective RTT, including the additional time due to delayed ACKs, or else it will retransmit unnecessarily. Thus, when delayed ACKs are in use, the receiver should reply with the TSval field from the earliest unacknowledged segment.

(B) A hole in the sequence space (segment(s) have been lost).

The sender will continue sending until the window is filled, and the receiver may be generating ACKs as these out-of-order segments arrive (e.g., to aid "fast retransmit").

The lost segment is probably a sign of congestion, and in that situation the sender should be conservative about retransmission. Furthermore, it is better to overestimate than underestimate the RTT. An ACK for an out-of-order segment should therefore contain the timestamp from the most recent segment that advanced the window.

The same situation occurs if segments are re-ordered by the network.

(C) A filled hole in the sequence space.

The segment that fills the hole represents the most recent measurement of the network characteristics. On the other hand, an RTT computed from an earlier segment would probably include the sender's retransmit time-out, badly biasing the sender's average RTT estimate. Thus, the timestamp from the latest segment (which filled the hole) must be echoed.

An algorithm that covers all three cases is described in the following rules for Timestamps option processing on a synchronized connection:

(1) The connection state is augmented with two 32-bit slots: TS.Recent holds a timestamp to be echoed in TSecr whenever a segment is sent, and Last.ACK.sent holds the ACK field from the last segment sent. Last.ACK.sent will equal RCV.NXT except when ACKs have been delayed.

RED HAT 1010
RED HAT V. CAS

(2) If Last.ACK.sent falls within the range of sequence numbers of an incoming segment:

SEG.SEQ <= Last.ACK.sent < SEG.SEQ + SEG.LEN

then the TSval from the segment is copied to TS.Recent; otherwise, the TSval is ignored.

(3) When a TSopt is sent, its TSecr field is set to the current TS.Recent value.

The following examples illustrate these rules. Here A, B, C... represent data segments occupying successive blocks of sequence numbers, and ACK(A),... represent the corresponding acknowledgment segments. Note that ACK(A) has the same sequence number as B. We show only one direction of timestamp echoing, for clarity.

o Packets arrive in sequence, and some of the ACKs are delayed.

By Case (A), the timestamp from the oldest unacknowledged segment is echoed.

```
                                          TS.Recent
        <A, TSval=1> ------------------->
                                              1
        <B, TSval=2> ------------------->
                                              1
        <C, TSval=3> ------------------->
                                              1
                 <---- <ACK(C), TSecr=1>
        (etc)
```

o Packets arrive out of order, and every packet is acknowledged.

By Case (B), the timestamp from the last segment that advanced the left window edge is echoed, until the missing segment arrives; it is echoed according to Case (C). The same sequence would occur if segments B and D were lost and retransmitted..

RED HAT 1010
RED HAT V. CAS

```
                                                    TS.Recent
      <A, TSval=1> -------------------->
                                                          1
               <---- <ACK(A), TSecr=1>
                                                          1
      <C, TSval=3> -------------------->
                                                          1
               <---- <ACK(A), TSecr=1>
                                                          1
      <B, TSval=2> -------------------->
                                                          2
               <---- <ACK(C), TSecr=2>
                                                          2
      <E, TSval=5> -------------------->
                                                          2
               <---- <ACK(C), TSecr=2>
                                                          2
      <D, TSval=4> -------------------->
                                                          4
               <---- <ACK(E), TSecr=4>
      (etc)
```

4. PAWS: PROTECT AGAINST WRAPPED SEQUENCE NUMBERS

4.1 Introduction

Section 4.2 describes a simple mechanism to reject old duplicate segments that might corrupt an open TCP connection; we call this mechanism PAWS (Protect Against Wrapped Sequence numbers). PAWS operates within a single TCP connection, using state that is saved in the connection control block. Section 4.3 and Appendix C discuss the implications of the PAWS mechanism for avoiding old duplicates from previous incarnations of the same connection.

4.2 The PAWS Mechanism

PAWS uses the same TCP Timestamps option as the RTTM mechanism described earlier, and assumes that every received TCP segment (including data and ACK segments) contains a timestamp SEG.TSval whose values are monotone non-decreasing in time. The basic idea is that a segment can be discarded as an old duplicate if it is received with a timestamp SEG.TSval less than some timestamp recently received on this connection.

In both the PAWS and the RTTM mechanism, the "timestamps" are 32-

RED HAT 1010
RED HAT V. CAS

bit unsigned integers in a modular 32-bit space. Thus, "less than" is defined the same way it is for TCP sequence numbers, and the same implementation techniques apply. If s and t are timestamp values, s < t if 0 < (t - s) < 2**31, computed in unsigned 32-bit arithmetic.

The choice of incoming timestamps to be saved for this comparison must guarantee a value that is monotone increasing. For example, we might save the timestamp from the segment that last advanced the left edge of the receive window, i.e., the most recent in-sequence segment. Instead, we choose the value TS.Recent introduced in Section 3.4 for the RTTM mechanism, since using a common value for both PAWS and RTTM simplifies the implementation of both. As Section 3.4 explained, TS.Recent differs from the timestamp from the last in-sequence segment only in the case of delayed ACKs, and therefore by less than one window. Either choice will therefore protect against sequence number wrap-around.

RTTM was specified in a symmetrical manner, so that TSval timestamps are carried in both data and ACK segments and are echoed in TSecr fields carried in returning ACK or data segments. PAWS submits all incoming segments to the same test, and therefore protects against duplicate ACK segments as well as data segments. (An alternative un-symmetric algorithm would protect against old duplicate ACKs: the sender of data would reject incoming ACK segments whose TSecr values were less than the TSecr saved from the last segment whose ACK field advanced the left edge of the send window. This algorithm was deemed to lack economy of mechanism and symmetry.)

TSval timestamps sent on {SYN} and {SYN,ACK} segments are used to initialize PAWS. PAWS protects against old duplicate non-SYN segments, and duplicate SYN segments received while there is a synchronized connection. Duplicate {SYN} and {SYN,ACK} segments received when there is no connection will be discarded by the normal 3-way handshake and sequence number checks of TCP.

It is recommended that RST segments NOT carry timestamps, and that RST segments be acceptable regardless of their timestamp. Old duplicate RST segments should be exceedingly unlikely, and their cleanup function should take precedence over timestamps.

4.2.1 Basic PAWS Algorithm

The PAWS algorithm requires the following processing to be performed on all incoming segments for a synchronized connection:

RED HAT 1010
RED HAT V. CAS

R1) If there is a Timestamps option in the arriving segment and SEG.TSval < TS.Recent and if TS.Recent is valid (see later discussion), then treat the arriving segment as not acceptable:

> Send an acknowledgement in reply as specified in RFC-793 page 69 and drop the segment.
>
> Note: it is necessary to send an ACK segment in order to retain TCP's mechanisms for detecting and recovering from half-open connections. For example, see Figure 10 of RFC-793.

R2) If the segment is outside the window, reject it (normal TCP processing)

R3) If an arriving segment satisfies: SEG.SEQ <= Last.ACK.sent (see Section 3.4), then record its timestamp in TS.Recent.

R4) If an arriving segment is in-sequence (i.e., at the left window edge), then accept it normally.

R5) Otherwise, treat the segment as a normal in-window, out-of-sequence TCP segment (e.g., queue it for later delivery to the user).

Steps R2, R4, and R5 are the normal TCP processing steps specified by RFC-793.

It is important to note that the timestamp is checked only when a segment first arrives at the receiver, regardless of whether it is in-sequence or it must be queued for later delivery. Consider the following example.

> Suppose the segment sequence: A.1, B.1, C.1, ..., Z.1 has been sent, where the letter indicates the sequence number and the digit represents the timestamp. Suppose also that segment B.1 has been lost. The timestamp in TS.TStamp is 1 (from A.1), so C.1, ..., Z.1 are considered acceptable and are queued. When B is retransmitted as segment B.2 (using the latest timestamp), it fills the hole and causes all the segments through Z to be acknowledged and passed to the user. The timestamps of the queued segments are *not* inspected again at this time, since they have already been accepted. When B.2 is accepted, TS.Stamp is set to 2.

This rule allows reasonable performance under loss. A full

RED HAT 1010
RED HAT V. CAS

window of data is in transit at all times, and after a loss a full window less one packet will show up out-of-sequence to be queued at the receiver (e.g., up to ~2**30 bytes of data); the timestamp option must not result in discarding this data.

In certain unlikely circumstances, the algorithm of rules R1-R4 could lead to discarding some segments unnecessarily, as shown in the following example:

> Suppose again that segments: A.1, B.1, C.1, ..., Z.1 have been sent in sequence and that segment B.1 has been lost. Furthermore, suppose delivery of some of C.1, ... Z.1 is delayed until AFTER the retransmission B.2 arrives at the receiver. These delayed segments will be discarded unnecessarily when they do arrive, since their timestamps are now out of date.

This case is very unlikely to occur. If the retransmission was triggered by a timeout, some of the segments C.1, ... Z.1 must have been delayed longer than the RTO time. This is presumably an unlikely event, or there would be many spurious timeouts and retransmissions. If B's retransmission was triggered by the "fast retransmit" algorithm, i.e., by duplicate ACKs, then the queued segments that caused these ACKs must have been received already.

Even if a segment were delayed past the RTO, the Fast Retransmit mechanism [Jacobson90c] will cause the delayed packets to be retransmitted at the same time as B.2, avoiding an extra RTT and therefore causing a very small performance penalty.

We know of no case with a significant probability of occurrence in which timestamps will cause performance degradation by unnecessarily discarding segments.

4.2.2 Timestamp Clock

It is important to understand that the PAWS algorithm does not require clock synchronization between sender and receiver. The sender's timestamp clock is used to stamp the segments, and the sender uses the echoed timestamp to measure RTT's. However, the receiver treats the timestamp as simply a monotone-increasing serial number, without any necessary connection to its clock. From the receiver's viewpoint, the timestamp is acting as a logical extension of the high-order bits of the sequence number.

RED HAT 1010
RED HAT V. CAS

The receiver algorithm does place some requirements on the frequency of the timestamp clock.

(a) The timestamp clock must not be "too slow".

It must tick at least once for each 2**31 bytes sent. In fact, in order to be useful to the sender for round trip timing, the clock should tick at least once per window's worth of data, and even with the RFC-1072 window extension, 2**31 bytes must be at least two windows.

To make this more quantitative, any clock faster than 1 tick/sec will reject old duplicate segments for link speeds of ~8 Gbps. A 1ms timestamp clock will work at link speeds up to 8 Tbps (8*10**12) bps!

(b) The timestamp clock must not be "too fast".

Its recycling time must be greater than MSL seconds. Since the clock (timestamp) is 32 bits and the worst-case MSL is 255 seconds, the maximum acceptable clock frequency is one tick every 59 ns.

However, it is desirable to establish a much longer recycle period, in order to handle outdated timestamps on idle connections (see Section 4.2.3), and to relax the MSL requirement for preventing sequence number wrap-around. With a 1 ms timestamp clock, the 32-bit timestamp will wrap its sign bit in 24.8 days. Thus, it will reject old duplicates on the same connection if MSL is 24.8 days or less. This appears to be a very safe figure; an MSL of 24.8 days or longer can probably be assumed by the gateway system without requiring precise MSL enforcement by the TTL value in the IP layer.

Based upon these considerations, we choose a timestamp clock frequency in the range 1 ms to 1 sec per tick. This range also matches the requirements of the RTTM mechanism, which does not need much more resolution than the granularity of the retransmit timer, e.g., tens or hundreds of milliseconds.

The PAWS mechanism also puts a strong monotonicity requirement on the sender's timestamp clock. The method of implementation of the timestamp clock to meet this requirement depends upon the system hardware and software.

* Some hosts have a hardware clock that is guaranteed to be monotonic between hardware resets.

RED HAT 1010
RED HAT V. CAS

* A clock interrupt may be used to simply increment a binary integer by 1 periodically.

* The timestamp clock may be derived from a system clock that is subject to being abruptly changed, by adding a variable offset value. This offset is initialized to zero. When a new timestamp clock value is needed, the offset can be adjusted as necessary to make the new value equal to or larger than the previous value (which was saved for this purpose).

4.2.3 Outdated Timestamps

If a connection remains idle long enough for the timestamp clock of the other TCP to wrap its sign bit, then the value saved in TS.Recent will become too old; as a result, the PAWS mechanism will cause all subsequent segments to be rejected, freezing the connection (until the timestamp clock wraps its sign bit again).

With the chosen range of timestamp clock frequencies (1 sec to 1 ms), the time to wrap the sign bit will be between 24.8 days and 24800 days. A TCP connection that is idle for more than 24 days and then comes to life is exceedingly unusual. However, it is undesirable in principle to place any limitation on TCP connection lifetimes.

We therefore require that an implementation of PAWS include a mechanism to "invalidate" the TS.Recent value when a connection is idle for more than 24 days. (An alternative solution to the problem of outdated timestamps would be to send keepalive segments at a very low rate, but still more often than the wrap-around time for timestamps, e.g., once a day. This would impose negligible overhead. However, the TCP specification has never included keepalives, so the solution based upon invalidation was chosen.)

Note that a TCP does not know the frequency, and therefore, the wraparound time, of the other TCP, so it must assume the worst. The validity of TS.Recent needs to be checked only if the basic PAWS timestamp check fails, i.e., only if SEG.TSval < TS.Recent. If TS.Recent is found to be invalid, then the segment is accepted, regardless of the failure of the timestamp check, and rule R3 updates TS.Recent with the TSval from the new segment.

To detect how long the connection has been idle, the TCP may

RED HAT 1010
RED HAT V. CAS

update a clock or timestamp value associated with the connection whenever TS.Recent is updated, for example. The details will be implementation-dependent.

4.2.4 Header Prediction

"Header prediction" [Jacobson90a] is a high-performance transport protocol implementation technique that is most important for high-speed links. This technique optimizes the code for the most common case, receiving a segment correctly and in order. Using header prediction, the receiver asks the question, "Is this segment the next in sequence?" This question can be answered in fewer machine instructions than the question, "Is this segment within the window?"

Adding header prediction to our timestamp procedure leads to the following recommended sequence for processing an arriving TCP segment:

H1) Check timestamp (same as step R1 above)

H2) Do header prediction: if segment is next in sequence and if there are no special conditions requiring additional processing, accept the segment, record its timestamp, and skip H3.

H3) Process the segment normally, as specified in RFC-793. This includes dropping segments that are outside the window and possibly sending acknowledgments, and queueing in-window, out-of-sequence segments.

Another possibility would be to interchange steps H1 and H2, i.e., to perform the header prediction step H2 FIRST, and perform H1 and H3 only when header prediction fails. This could be a performance improvement, since the timestamp check in step H1 is very unlikely to fail, and it requires interval arithmetic on a finite field, a relatively expensive operation. To perform this check on every single segment is contrary to the philosophy of header prediction. We believe that this change might reduce CPU time for TCP protocol processing by up to 5-10% on high-speed networks.

However, putting H2 first would create a hazard: a segment from 2**32 bytes in the past might arrive at exactly the wrong time and be accepted mistakenly by the header-prediction step. The following reasoning has been introduced [Jacobson90b] to show that the probability of this failure is negligible.

RED HAT 1010
RED HAT V. CAS

If all segments are equally likely to show up as old duplicates, then the probability of an old duplicate exactly matching the left window edge is the maximum segment size (MSS) divided by the size of the sequence space. This ratio must be less than 2**-16, since MSS must be < 2**16; for example, it will be (2**12)/(2**32) = 2**-20 for an FDDI link. However, the older a segment is, the less likely it is to be retained in the Internet, and under any reasonable model of segment lifetime the probability of an old duplicate exactly at the left window edge must be much smaller than 2**-16.

The 16 bit TCP checksum also allows a basic unreliability of one part in 2**16. A protocol mechanism whose reliability exceeds the reliability of the TCP checksum should be considered "good enough", i.e., it won't contribute significantly to the overall error rate. We therefore believe we can ignore the problem of an old duplicate being accepted by doing header prediction before checking the timestamp.

However, this probabilistic argument is not universally accepted, and the consensus at present is that the performance gain does not justify the hazard in the general case. It is therefore recommended that H2 follow H1.

4.3. Duplicates from Earlier Incarnations of Connection

The PAWS mechanism protects against errors due to sequence number wrap-around on high-speed connection. Segments from an earlier incarnation of the same connection are also a potential cause of old duplicate errors. In both cases, the TCP mechanisms to prevent such errors depend upon the enforcement of a maximum segment lifetime (MSL) by the Internet (IP) layer (see Appendix of RFC-1185 for a detailed discussion). Unlike the case of sequence space wrap-around, the MSL required to prevent old duplicate errors from earlier incarnations does not depend upon the transfer rate. If the IP layer enforces the recommended 2 minute MSL of TCP, and if the TCP rules are followed, TCP connections will be safe from earlier incarnations, no matter how high the network speed. Thus, the PAWS mechanism is not required for this case.

We may still ask whether the PAWS mechanism can provide additional security against old duplicates from earlier connections, allowing us to relax the enforcement of MSL by the IP layer. Appendix B explores this question, showing that further assumptions and/or mechanisms are required, beyond those of PAWS. This is not part of the current extension.

RED HAT 1010
RED HAT V. CAS

5. CONCLUSIONS AND ACKNOWLEDGMENTS

This memo presented a set of extensions to TCP to provide efficient operation over large-bandwidth*delay-product paths and reliable operation over very high-speed paths. These extensions are designed to provide compatible interworking with TCP's that do not implement the extensions.

These mechanisms are implemented using new TCP options for scaled windows and timestamps. The timestamps are used for two distinct mechanisms: RTTM (Round Trip Time Measurement) and PAWS (Protect Against Wrapped Sequences).

The Window Scale option was originally suggested by Mike St. Johns of USAF/DCA. The present form of the option was suggested by Mike Karels of UC Berkeley in response to a more cumbersome scheme defined by Van Jacobson. Lixia Zhang helped formulate the PAWS mechanism description in RFC-1185.

Finally, much of this work originated as the result of discussions within the End-to-End Task Force on the theoretical limitations of transport protocols in general and TCP in particular. More recently, task force members and other on the end2end-interest list have made valuable contributions by pointing out flaws in the algorithms and the documentation. The authors are grateful for all these contributions.

6. REFERENCES

[Clark87] Clark, D., Lambert, M., and L. Zhang, "NETBLT: A Bulk Data Transfer Protocol", RFC 998, MIT, March 1987.

[Garlick77] Garlick, L., R. Rom, and J. Postel, "Issues in Reliable Host-to-Host Protocols", Proc. Second Berkeley Workshop on Distributed Data Management and Computer Networks, May 1977.

[Hamming77] Hamming, R., "Digital Filters", ISBN 0-13-212571-4, Prentice Hall, Englewood Cliffs, N.J., 1977.

[Cheriton88] Cheriton, D., "VMTP: Versatile Message Transaction Protocol", RFC 1045, Stanford University, February 1988.

[Jacobson88a] Jacobson, V., "Congestion Avoidance and Control", SIGCOMM '88, Stanford, CA., August 1988.

[Jacobson88b] Jacobson, V., and R. Braden, "TCP Extensions for Long-Delay Paths", RFC-1072, LBL and USC/Information Sciences Institute, October 1988.

RED HAT 1010
RED HAT V. CAS

[Jacobson90a] Jacobson, V., "4BSD Header Prediction", ACM Computer Communication Review, April 1990.

[Jacobson90b] Jacobson, V., Braden, R., and Zhang, L., "TCP Extension for High-Speed Paths", RFC-1185, LBL and USC/Information Sciences Institute, October 1990.

[Jacobson90c] Jacobson, V., "Modified TCP congestion avoidance algorithm", Message to end2end-interest mailing list, April 1990.

[Jain86] Jain, R., "Divergence of Timeout Algorithms for Packet Retransmissions", Proc. Fifth Phoenix Conf. on Comp. and Comm., Scottsdale, Arizona, March 1986.

[Karn87] Karn, P. and C. Partridge, "Estimating Round-Trip Times in Reliable Transport Protocols", Proc. SIGCOMM '87, Stowe, VT, August 1987.

[McKenzie89] McKenzie, A., "A Problem with the TCP Big Window Option", RFC 1110, BBN STC, August 1989.

[Nagle84] Nagle, J., "Congestion Control in IP/TCP Internetworks", RFC 896, FACC, January 1984.

[NBS85] Colella, R., Aronoff, R., and K. Mills, "Performance Improvements for ISO Transport", Ninth Data Comm Symposium, published in ACM SIGCOMM Comp Comm Review, vol. 15, no. 5, September 1985.

[Postel81] Postel, J., "Transmission Control Protocol - DARPA Internet Program Protocol Specification", RFC 793, DARPA, September 1981.

[Velten84] Velten, D., Hinden, R., and J. Sax, "Reliable Data Protocol", RFC 908, BBN, July 1984.

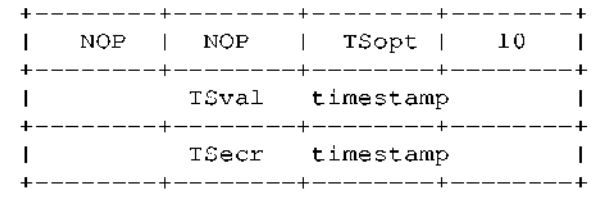
[Watson81] Watson, R., "Timer-based Mechanisms in Reliable Transport Protocol Connection Management", Computer Networks, Vol. 5, 1981.

[Zhang86] Zhang, L., "Why TCP Timers Don't Work Well", Proc. SIGCOMM '86, Stowe, Vt., August 1986.

RED HAT 1010
RED HAT V. CAS

APPENDIX A: IMPLEMENTATION SUGGESTIONS

The following layouts are recommended for sending options on non-SYN segments, to achieve maximum feasible alignment of 32-bit and 64-bit machines.



APPENDIX B: DUPLICATES FROM EARLIER CONNECTION INCARNATIONS

There are two cases to be considered: (1) a system crashing (and losing connection state) and restarting, and (2) the same connection being closed and reopened without a loss of host state. These will be described in the following two sections.

B.1 System Crash with Loss of State

TCP's quiet time of one MSL upon system startup handles the loss of connection state in a system crash/restart. For an explanation, see for example "When to Keep Quiet" in the TCP protocol specification [Postel81]. The MSL that is required here does not depend upon the transfer speed. The current TCP MSL of 2 minutes seems acceptable as an operational compromise, as many host systems take this long to boot after a crash.

However, the timestamp option may be used to ease the MSL requirements (or to provide additional security against data corruption). If timestamps are being used and if the timestamp clock can be guaranteed to be monotonic over a system crash/restart, i.e., if the first value of the sender's timestamp clock after a crash/restart can be guaranteed to be greater than the last value before the restart, then a quiet time will be unnecessary.

To dispense totally with the quiet time would require that the host clock be synchronized to a time source that is stable over the crash/restart period, with an accuracy of one timestamp clock tick or better. We can back off from this strict requirement to take advantage of approximate clock synchronization. Suppose that the clock is always re-synchronized to within N timestamp clock

RED HAT 1010
RED HAT V. CAS

ticks and that booting (extended with a quiet time, if necessary) takes more than N ticks. This will guarantee monotonicity of the timestamps, which can then be used to reject old duplicates even without an enforced MSL.

B.2 Closing and Reopening a Connection

When a TCP connection is closed, a delay of 2*MSL in TIME-WAIT state ties up the socket pair for 4 minutes (see Section 3.5 of [Postel81]. Applications built upon TCP that close one connection and open a new one (e.g., an FTP data transfer connection using Stream mode) must choose a new socket pair each time. The TIME-WAIT delay serves two different purposes:

(a) Implement the full-duplex reliable close handshake of TCP.

The proper time to delay the final close step is not really related to the MSL; it depends instead upon the RTO for the FIN segments and therefore upon the RTT of the path. (It could be argued that the side that is sending a FIN knows what degree of reliability it needs, and therefore it should be able to determine the length of the TIME-WAIT delay for the FIN's recipient. This could be accomplished with an appropriate TCP option in FIN segments.)

Although there is no formal upper-bound on RTT, common network engineering practice makes an RTT greater than 1 minute very unlikely. Thus, the 4 minute delay in TIME-WAIT state works satisfactorily to provide a reliable full-duplex TCP close. Note again that this is independent of MSL enforcement and network speed.

The TIME-WAIT state could cause an indirect performance problem if an application needed to repeatedly close one connection and open another at a very high frequency, since the number of available TCP ports on a host is less than 2**16. However, high network speeds are not the major contributor to this problem; the RTT is the limiting factor in how quickly connections can be opened and closed. Therefore, this problem will be no worse at high transfer speeds.

(b) Allow old duplicate segments to expire.

To replace this function of TIME-WAIT state, a mechanism would have to operate across connections. PAWS is defined strictly within a single connection; the last timestamp is TS.Recent is kept in the connection control block, and

discarded when a connection is closed.

An additional mechanism could be added to the TCP, a per-host cache of the last timestamp received from any connection. This value could then be used in the PAWS mechanism to reject old duplicate segments from earlier incarnations of the connection, if the timestamp clock can be guaranteed to have ticked at least once since the old connection was open. This would require that the TIME-WAIT delay plus the RTT together must be at least one tick of the sender's timestamp clock. Such an extension is not part of the proposal of this RFC.

Note that this is a variant on the mechanism proposed by Garlick, Rom, and Postel [Garlick77], which required each host to maintain connection records containing the highest sequence numbers on every connection. Using timestamps instead, it is only necessary to keep one quantity per remote host, regardless of the number of simultaneous connections to that host.

APPENDIX C: CHANGES FROM RFC-1072, RFC-1185

The protocol extensions defined in this document differ in several important ways from those defined in RFC-1072 and RFC-1185.

(a) SACK has been deferred to a later memo.

(b) The detailed rules for sending timestamp replies (see Section 3.4) differ in important ways. The earlier rules could result in an under-estimate of the RTT in certain cases (packets dropped or out of order).

(c) The same value TS.Recent is now shared by the two distinct mechanisms RTTM and PAWS. This simplification became possible because of change (b).

(d) An ambiguity in RFC-1185 was resolved in favor of putting timestamps on ACK as well as data segments. This supports the symmetry of the underlying TCP protocol.

(e) The echo and echo reply options of RFC-1072 were combined into a single Timestamps option, to reflect the symmetry and to simplify processing.

(f) The problem of outdated timestamps on long-idle connections, discussed in Section 4.2.2, was realized and resolved.

(g) RFC-1185 recommended that header prediction take precedence over the timestamp check. Based upon some scepticism about the probabilistic arguments given in Section 4.2.4, it was decided to recommend that the timestamp check be performed first.

(h) The spec was modified so that the extended options will be sent on <SYN,ACK> segments only when they are received in the corresponding <SYN> segments. This provides the most conservative possible conditions for interoperation with implementations without the extensions.

In addition to these substantive changes, the present RFC attempts to specify the algorithms unambiguously by presenting modifications to the Event Processing rules of RFC-793; see Appendix E.

RED HAT 1010
RED HAT V. CAS

APPENDIX D: SUMMARY OF NOTATION

The following notation has been used in this document.

Options

WSopt:	TCP Window Scale Option
TSopt:	TCP Timestamps Option

Option Fields

shift.cnt:	Window scale byte in WSopt.
TSval:	32-bit Timestamp Value field in TSopt.
TSecr:	32-bit Timestamp Reply field in TSopt.

Option Fields in Current Segment

SEG.TSval:	TSval field from TSopt in current segment.
SEG.TSecr:	TSecr field from TSopt in current segment.
SEG.WSopt:	8-bit value in WSopt

Clock Values

my.TSclock:	Local source of 32-bit timestamp values
my.TSclock.rate:	Period of my.TSclock (1 ms to 1 sec).

Per-Connection State Variables

TS.Recent:	Latest received Timestamp
Last.ACK.sent:	Last ACK field sent
Snd.TS.OK:	1-bit flag
Snd.WS.OK:	1-bit flag
Rcv.Wind.Scale:	Receive window scale power
Snd.Wind.Scale:	Send window scale power

APPENDIX E: EVENT PROCESSING

Event Processing

OPEN Call

...
An initial send sequence number (ISS) is selected. Send a SYN segment of the form:

<SEQ=ISS><CTL=SYN><TSval=my.TSclock><WSopt=Rcv.Wind.Scale>

...

SEND Call

CLOSED STATE (i.e., TCB does not exist)

...

LISTEN STATE

If the foreign socket is specified, then change the connection from passive to active, select an ISS. Send a SYN segment containing the options: <TSval=my.TSclock> and <WSopt=Rcv.Wind.Scale>. Set SND.UNA to ISS, SND.NXT to ISS+1. Enter SYN-SENT state. ...

SYN-SENT STATE
SYN-RECEIVED STATE

...

ESTABLISHED STATE
CLOSE-WAIT STATE

Segmentize the buffer and send it with a piggybacked acknowledgment (acknowledgment value = RCV.NXT). ...

If the urgent flag is set ...

If the Snd.TS.OK flag is set, then include the TCP Timestamps option <TSval=my.TSclock,TSecr=TS.Recent> in each data segment.

Scale the receive window for transmission in the segment header:

SEG.WND = (SND.WND >> Rcv.Wind.Scale).

```
  SEGMENT ARRIVES

     ...

    If the state is LISTEN then

      first check for an RST

        ...

      second check for an ACK

        ...

      third check for a SYN

        if the SYN bit is set, check the security.  If the ...

         ...

        If the SEG.PRC is less than the TCB.PRC then continue.

        Check for a Window Scale option (WSopt); if one is found, save
        SEG.WSopt in Snd.Wind.Scale and set Snd.WS.OK flag on.
        Otherwise, set both Snd.Wind.Scale and Rcv.Wind.Scale to zero
        and clear Snd.WS.OK flag.

        Check for a TSopt option; if one is found, save SEG.TSval in the
        variable TS.Recent and turn on the Snd.TS.OK bit.

        Set RCV.NXT to SEG.SEQ+1, IRS is set to SEG.SEQ and any other
        control or text should be queued for processing later.  ISS
        should be selected and a SYN segment sent of the form:

          <SEQ=ISS><ACK=RCV.NXT><CTL=SYN,ACK>

        If the Snd.WS.OK bit is on, include a WSopt option
        <WSopt=Rcv.Wind.Scale> in this segment.  If the Snd.TS.OK bit is
        on, include a TSopt <TSval=my.TSclock,TSecr=TS.Recent> in this
        segment.  Last.ACK.sent is set to RCV.NXT.

        SND.NXT is set to ISS+1 and SND.UNA to ISS.  The connection
        state should be changed to SYN-RECEIVED.  Note that any other
        incoming control or data (combined with SYN) will be processed
        in the SYN-RECEIVED state, but processing of SYN and ACK should
        not be repeated.  If the listen was not fully specified (i.e.,
        the foreign socket was not fully specified), then the
        unspecified fields should be filled in now.
```


RED HAT 1010
RED HAT V. CAS

```
      fourth other text or control

       ...

    If the state is SYN-SENT then

      first check the ACK bit

        ...

      fourth check the SYN bit

         ...

        If the SYN bit is on and the security/compartment and precedence
        are acceptable then, RCV.NXT is set to SEG.SEQ+1, IRS is set to
        SEG.SEQ, and any acknowledgements on the retransmission queue
        which are thereby acknowledged should be removed.

        Check for a Window Scale option (WSopt); if is found, save
        SEG.WSopt in Snd.Wind.Scale; otherwise, set both Snd.Wind.Scale
        and Rcv.Wind.Scale to zero.

        Check for a TSopt option; if one is found, save SEG.TSval in
        variable TS.Recent and turn on the Snd.TS.OK bit in the
        connection control block.  If the ACK bit is set, use my.TSclock
        - SEG.TSecr as the initial RTT estimate.

        If SND.UNA > ISS (our SYN has been ACKed), change the connection
        state to ESTABLISHED, form an ACK segment:

            <SEQ=SND.NXT><ACK=RCV.NXT><CTL=ACK>

        and send it.  If the Snd.Echo.OK bit is on, include a TSopt
        option <TSval=my.TSclock,TSecr=TS.Recent> in this ACK segment.
        Last.ACK.sent is set to RCV.NXT.

        Data or controls which were queued for transmission may be
        included.  If there are other controls or text in the segment
        then continue processing at the sixth step below where the URG
        bit is checked, otherwise return.

        Otherwise enter SYN-RECEIVED, form a SYN,ACK segment:

            <SEQ=ISS><ACK=RCV.NXT><CTL=SYN,ACK>

        and send it.  If the Snd.Echo.OK bit is on, include a TSopt
        option <TSval=my.TSclock,TSecr=TS.Recent> in this segment.  If
```

the Snd.WS.OK bit is on, include a WSopt option <WSopt=Rcv.Wind.Scale> in this segment. Last.ACK.sent is set to RCV.NXT.

If there are other controls or text in the segment, queue them for processing after the ESTABLISHED state has been reached, return.

fifth, if neither of the SYN or RST bits is set then drop the segment and return.

Otherwise,

First, check sequence number

SYN-RECEIVED STATE
ESTABLISHED STATE
FIN-WAIT-1 STATE
FIN-WAIT-2 STATE
CLOSE-WAIT STATE
CLOSING STATE
LAST-ACK STATE
TIME-WAIT STATE

Segments are processed in sequence. Initial tests on arrival are used to discard old duplicates, but further processing is done in SEG.SEQ order. If a segment's contents straddle the boundary between old and new, only the new parts should be processed.

Rescale the received window field:

```
TrueWindow = SEG.WND << Snd.Wind.Scale,
```

and use "TrueWindow" in place of SEG.WND in the following steps.

Check whether the segment contains a Timestamps option and bit Snd.TS.OK is on. If so:

If SEG.TSval < TS.Recent, then test whether connection has been idle less than 24 days; if both are true, then the segment is not acceptable; follow steps below for an unacceptable segment.

If SEG.SEQ is equal to Last.ACK.sent, then save SEG.ECopt in variable TS.Recent.

RED HAT 1010
RED HAT V. CAS

There are four cases for the acceptability test for an incoming segment:

...

If an incoming segment is not acceptable, an acknowledgment should be sent in reply (unless the RST bit is set, if so drop the segment and return):

<SEQ=SND.NXT><ACK=RCV.NXT><CTL=ACK>

Last.ACK.sent is set to SEG.ACK of the acknowledgment. If the Snd.Echo.OK bit is on, include the Timestamps option <TSval=my.TSclock,TSecr=TS.Recent> in this ACK segment. Set Last.ACK.sent to SEG.ACK and send the ACK segment. After sending the acknowledgment, drop the unacceptable segment and return.

...

fifth check the ACK field.

if the ACK bit is off drop the segment and return.

if the ACK bit is on

...

ESTABLISHED STATE

If SND.UNA < SEG.ACK =< SND.NXT then, set SND.UNA <- SEG.ACK. Also compute a new estimate of round-trip time. If Snd.TS.OK bit is on, use my.TSclock - SEG.TSecr; otherwise use the elapsed time since the first segment in the retransmission queue was sent. Any segments on the retransmission queue which are thereby entirely acknowledged...

...

Seventh, process the segment text.

ESTABLISHED STATE
FIN-WAIT-1 STATE
FIN-WAIT-2 STATE

...

Send an acknowledgment of the form:

RED HAT 1010
RED HAT V. CAS

```
<SEQ=SND.NXT><ACK=RCV.NXT><CTL=ACK>
```

If the Snd.TS.OK bit is on, include Timestamps option <TSval=my.TSclock,TSecr=TS.Recent> in this ACK segment. Set Last.ACK.sent to SEG.ACK of the acknowledgment, and send it. This acknowledgment should be piggy-backed on a segment being transmitted if possible without incurring undue delay.

...

Security Considerations

Security issues are not discussed in this memo.

Authors' Addresses

Van Jacobson
University of California
Lawrence Berkeley Laboratory
Mail Stop 46A
Berkeley, CA 94720

Phone: (415) 486-6411
EMail: van@CSAM.LBL.GOV

Bob Braden
University of Southern California
Information Sciences Institute
4676 Admiralty Way
Marina del Rey, CA 90292

Phone: (310) 822-1511
EMail: Braden@ISI.EDU

Dave Borman
Cray Research
655-E Lone Oak Drive
Eagan, MN 55121

Phone: (612) 683-5571
Email: dab@cray.com

Network Working Group
Request for Comments: 2002
Category: Standards Track

C. Perkins, Editor
IBM

October 1996

IP Mobility Support

Status of this Memo

This document specifies an Internet standards track protocol for the Internet community, and requests discussion and suggestions for improvements. Please refer to the current edition of the "Internet Official Protocol Standards" (STD 1) for the standardization state and status of this protocol. Distribution of this memo is unlimited.

Abstract

This document specifies protocol enhancements that allow transparent routing of IP datagrams to mobile nodes in the Internet. Each mobile node is always identified by its home address, regardless of its current point of attachment to the Internet. While situated away from its home, a mobile node is also associated with a care-of address, which provides information about its current point of attachment to the Internet. The protocol provides for registering the care-of address with a home agent. The home agent sends datagrams destined for the mobile node through a tunnel to the care-of address. After arriving at the end of the tunnel, each datagram is then delivered to the mobile node.

Table of Contents

1. Introduction 3
    1.1. Protocol Requirements 3
    1.2. Goals 4
    1.3. Assumptions 4
    1.4. Applicability 4
    1.5. New Architectural Entities 5
    1.6. Terminology 6
    1.7. Protocol Overview 8
    1.8. Specification Language 11
    1.9. Message Format and Protocol Extensibility 12
2. Agent Discovery 14
    2.1. Agent Advertisement 14
        2.1.1. Mobility Agent Advertisement Extension 16
        2.1.2. Prefix-Lengths Extension 18
        2.1.3. One-byte Padding Extension 19
    2.2. Agent Solicitation 19
    2.3. Foreign Agent and Home Agent Considerations 19
        2.3.1. Advertised Router Addresses 20

RED HAT 1010
RED HAT V. CAS

2.3.2. Sequence Numbers and Rollover Handling 21
2.4. Mobile Node Considerations 21
2.4.1. Registration Required 22
2.4.2. Move Detection 22
2.4.3. Returning Home 24
2.4.4. Sequence Numbers and Rollover Handling 24
3. Registration 24
3.1. Registration Overview 25
3.2. Authentication 26
3.3. Registration Request 26
3.4. Registration Reply 29
3.5. Registration Extensions 32
3.5.1. Computing Authentication Extension Values 32
3.5.2. Mobile-Home Authentication Extension 33
3.5.3. Mobile-Foreign Authentication Extension 33
3.5.4. Foreign-Home Authentication Extension 34
3.6. Mobile Node Considerations 34
3.6.1. Sending Registration Requests 36
3.6.2. Receiving Registration Replies 40
3.6.3. Registration Retransmission 42
3.7. Foreign Agent Considerations 43
3.7.1. Configuration and Registration Tables 44
3.7.2. Receiving Registration Requests 44
3.7.3. Receiving Registration Replies 47
3.8. Home Agent Considerations 49
3.8.1. Configuration and Registration Tables 49
3.8.2. Receiving Registration Requests 49
3.8.3. Sending Registration Replies 53
4. Routing Considerations 55
4.1. Encapsulation Types 56
4.2. Unicast Datagram Routing 56
4.2.1. Mobile Node Considerations 56
4.2.2. Foreign Agent Considerations 57
4.2.3. Home Agent Considerations 58
4.3. Broadcast Datagrams 59
4.4. Multicast Datagram Routing 60
4.5. Mobile Routers 61
4.6. ARP, Proxy ARP, and Gratuitous ARP 62
5. Security Considerations 66
5.1. Message Authentication Codes 66
5.2. Areas of Security Concern in this Protocol 66
5.3. Key Management 67
5.4. Picking Good Random Numbers 67
5.5. Privacy 67
5.6. Replay Protection for Registration Requests 68
5.6.1. Replay Protection using Timestamps 68
5.6.2. Replay Protection using Nonces 69
6. Acknowledgments 71

RED HAT 1010
RED HAT V. CAS

A. Patent Issues 72
A.1. IBM Patent #5,159,592 72
A.2. IBM Patent #5,148,479 72
B. Link-Layer Considerations 73
C. TCP Considerations 73
C.1. TCP Timers 73
C.2. TCP Congestion Management 73
D. Example Scenarios 74
D.1. Registering with a Foreign Agent Care-of Address 74
D.2. Registering with a Co-Located Care-of Address 75
D.3. Deregistration 76
E. Applicability of Prefix Lengths Extension 76
Editor's Address 79

1. Introduction

IP version 4 assumes that a node's IP address uniquely identifies the node's point of attachment to the Internet. Therefore, a node must be located on the network indicated by its IP address in order to receive datagrams destined to it; otherwise, datagrams destined to the node would be undeliverable. For a node to change its point of attachment without losing its ability to communicate, currently one of the two following mechanisms must typically be employed:

a) the node must change its IP address whenever it changes its point of attachment, or

b) host-specific routes must be propagated throughout much of the Internet routing fabric.

Both of these alternatives are often unacceptable. The first makes it impossible for a node to maintain transport and higher-layer connections when the node changes location. The second has obvious and severe scaling problems, especially relevant considering the explosive growth in sales of notebook (mobile) computers.

A new, scalable, mechanism is required for accommodating node mobility within the Internet. This document defines such a mechanism, which enables nodes to change their point of attachment to the Internet without changing their IP address.

1.1. Protocol Requirements

A mobile node must be able to communicate with other nodes after changing its link-layer point of attachment to the Internet, yet without changing its IP address.

RED HAT 1010
RED HAT V. CAS

A mobile node must be able to communicate with other nodes that do not implement these mobility functions. No protocol enhancements are required in hosts or routers that are not acting as any of the new architectural entities introduced in Section 1.5.

All messages used to update another node as to the location of a mobile node must be authenticated in order to protect against remote redirection attacks.

1.2. Goals

The link by which a mobile node is directly attached to the Internet may often be a wireless link. This link may thus have a substantially lower bandwidth and higher error rate than traditional wired networks. Moreover, mobile nodes are likely to be battery powered, and minimizing power consumption is important. Therefore, the number of administrative messages sent over the link by which a mobile node is directly attached to the Internet should be minimized, and the size of these messages should be kept as small as is reasonably possible.

1.3. Assumptions

The protocols defined in this document place no additional constraints on the assignment of IP addresses. That is, a mobile node can be assigned an IP address by the organization that owns the machine.

This protocol assumes that mobile nodes will generally not change their point of attachment to the Internet more frequently than once per second.

This protocol assumes that IP unicast datagrams are routed based on the destination address in the datagram header (and not, for example, by source address).

1.4. Applicability

Mobile IP is intended to enable nodes to move from one IP subnet to another. It is just as suitable for mobility across homogeneous media as it is for mobility across heterogeneous media. That is, Mobile IP facilitates node movement from one Ethernet segment to another as well as it accommodates node movement from an Ethernet segment to a wireless LAN, as long as the mobile node's IP address remains the same after such a movement.

One can think of Mobile IP as solving the "macro" mobility management problem. It is less well suited for more "micro" mobility management

RED HAT 1010
RED HAT V. CAS

applications -- for example, handoff amongst wireless transceivers, each of which covers only a very small geographic area. As long as node movement does not occur between points of attachment on different IP subnets, link-layer mechanisms for mobility (i.e., link-layer handoff) may offer faster convergence and far less overhead than Mobile IP.

1.5. New Architectural Entities

Mobile IP introduces the following new functional entities:

Mobile Node

A host or router that changes its point of attachment from one network or subnetwork to another. A mobile node may change its location without changing its IP address; it may continue to communicate with other Internet nodes at any location using its (constant) IP address, assuming link-layer connectivity to a point of attachment is available.

Home Agent

A router on a mobile node's home network which tunnels datagrams for delivery to the mobile node when it is away from home, and maintains current location information for the mobile node.

Foreign Agent

A router on a mobile node's visited network which provides routing services to the mobile node while registered. The foreign agent detunnels and delivers datagrams to the mobile node that were tunneled by the mobile node's home agent. For datagrams sent by a mobile node, the foreign agent may serve as a default router for registered mobile nodes.

A mobile node is given a long-term IP address on a home network. This home address is administered in the same way as a "permanent" IP address is provided to a stationary host. When away from its home network, a "care-of address" is associated with the mobile node and reflects the mobile node's current point of attachment. The mobile node uses its home address as the source address of all IP datagrams that it sends, except where otherwise described in this document for datagrams sent for certain mobility management functions (e.g., as in Section 3.6.1.1).

RED HAT 1010
RED HAT V. CAS

1.6. Terminology

This document frequently uses the following terms:

Agent Advertisement
: An advertisement message constructed by attaching a special Extension to a router advertisement [4] message.

Care-of Address
: The termination point of a tunnel toward a mobile node, for datagrams forwarded to the mobile node while it is away from home. The protocol can use two different types of care-of address: a "foreign agent care-of address" is an address of a foreign agent with which the mobile node is registered, and a "co-located care-of address" is an externally obtained local address which the mobile node has associated with one of its own network interfaces.

Correspondent Node
: A peer with which a mobile node is communicating. A correspondent node may be either mobile or stationary.

Foreign Network
: Any network other than the mobile node's Home Network.

Home Address
: An IP address that is assigned for an extended period of time to a mobile node. It remains unchanged regardless of where the node is attached to the Internet.

Home Network
: A network, possibly virtual, having a network prefix matching that of a mobile node's home address. Note that standard IP routing mechanisms will deliver datagrams destined to a mobile node's Home Address to the mobile node's Home Network.

Link
: A facility or medium over which nodes can communicate at the link layer. A link underlies the network layer.

Link-Layer Address
: The address used to identify an endpoint of some communication over a physical link. Typically, the Link-Layer address is an interface's Media Access Control (MAC) address.

Mobility Agent
: Either a home agent or a foreign agent.

RED HAT 1010
RED HAT V. CAS

Mobility Binding
: The association of a home address with a care-of address, along with the remaining lifetime of that association.

Mobility Security Association
: A collection of security contexts, between a pair of nodes, which may be applied to Mobile IP protocol messages exchanged between them. Each context indicates an authentication algorithm and mode (Section 5.1), a secret (a shared key, or appropriate public/private key pair), and a style of replay protection in use (Section 5.6).

Node
: A host or a router.

Nonce
: A randomly chosen value, different from previous choices, inserted in a message to protect against replays.

Security Parameter Index (SPI)
: An index identifying a security context between a pair of nodes among the contexts available in the Mobility Security Association. SPI values 0 through 255 are reserved and MUST NOT be used in any Mobility Security Association.

Tunnel
: The path followed by a datagram while it is encapsulated. The model is that, while it is encapsulated, a datagram is routed to a knowledgeable decapsulating agent, which decapsulates the datagram and then correctly delivers it to its ultimate destination.

Virtual Network
: A network with no physical instantiation beyond a router (with a physical network interface on another network). The router (e.g., a home agent) generally advertises reachability to the virtual network using conventional routing protocols.

Visited Network
: A network other than a mobile node's Home Network, to which the mobile node is currently connected.

Visitor List
: The list of mobile nodes visiting a foreign agent.

RED HAT 1010
RED HAT V. CAS
339

1.7. Protocol Overview

The following support services are defined for Mobile IP:

Agent Discovery
: Home agents and foreign agents may advertise their availability on each link for which they provide service. A newly arrived mobile node can send a solicitation on the link to learn if any prospective agents are present.

Registration
: When the mobile node is away from home, it registers its care-of address with its home agent. Depending on its method of attachment, the mobile node will register either directly with its home agent, or through a foreign agent which forwards the registration to the home agent.

The following steps provide a rough outline of operation of the Mobile IP protocol:

- Mobility agents (i.e., foreign agents and home agents) advertise their presence via Agent Advertisement messages (Section 2). A mobile node may optionally solicit an Agent Advertisement message from any locally attached mobility agents through an Agent Solicitation message.

- A mobile node receives these Agent Advertisements and determines whether it is on its home network or a foreign network.

- When the mobile node detects that it is located on its home network, it operates without mobility services. If returning to its home network from being registered elsewhere, the mobile node deregisters with its home agent, through exchange of a Registration Request and Registration Reply message with it.

- When a mobile node detects that it has moved to a foreign network, it obtains a care-of address on the foreign network. The care-of address can either be determined from a foreign agent's advertisements (a foreign agent care-of address), or by some external assignment mechanism such as DHCP [6] (a co-located care-of address).

- The mobile node operating away from home then registers its new care-of address with its home agent through exchange of a Registration Request and Registration Reply message with it, possibly via a foreign agent (Section 3).

RED HAT 1010
RED HAT V. CAS

- Datagrams sent to the mobile node's home address are intercepted by its home agent, tunneled by the home agent to the mobile node's care-of address, received at the tunnel endpoint (either at a foreign agent or at the mobile node itself), and finally delivered to the mobile node (Section 4.2.3).

- In the reverse direction, datagrams sent by the mobile node are generally delivered to their destination using standard IP routing mechanisms, not necessarily passing through the home agent.

When away from home, Mobile IP uses protocol tunneling to hide a mobile node's home address from intervening routers between its home network and its current location. The tunnel terminates at the mobile node's care-of address. The care-of address must be an address to which datagrams can be delivered via conventional IP routing. At the care-of address, the original datagram is removed from the tunnel and delivered to the mobile node.

Mobile IP provides two alternative modes for the acquisition of a care-of address:

- A "foreign agent care-of address" is a care-of address provided by a foreign agent through its Agent Advertisement messages. In this case, the care-of address is an IP address of the foreign agent. In this mode, the foreign agent is the endpoint of the tunnel and, upon receiving tunneled datagrams, decapsulates them and delivers the inner datagram to the mobile node. This mode of acquisition is preferred because it allows many mobile nodes to share the same care-of address and therefore does not place unnecessary demands on the already limited IPv4 address space.

- A "co-located care-of address" is a care-of address acquired by the mobile node as a local IP address through some external means, which the mobile node then associates with one of its own network interfaces. The address may be dynamically acquired as a temporary address by the mobile node such as through DHCP [6], or may be owned by the mobile node as a long-term address for its use only while visiting some foreign network. Specific external methods of acquiring a local IP address for use as a co-located care-of address are beyond the scope of this document. When using a co-located care-of address, the mobile node serves as the endpoint of the tunnel and itself performs decapsulation of the datagrams tunneled to it.

The mode of using a co-located care-of address has the advantage that it allows a mobile node to function without a foreign agent, for example, in networks that have not yet deployed a foreign agent.

RED HAT 1010
RED HAT V. CAS

It does, however, place additional burden on the IPv4 address space because it requires a pool of addresses within the foreign network to be made available to visiting mobile nodes. It is difficult to efficiently maintain pools of addresses for each subnet that may permit mobile nodes to visit.

It is important to understand the distinction between the care-of address and the foreign agent functions. The care-of address is simply the endpoint of the tunnel. It might indeed be an address of a foreign agent (a foreign agent care-of address), but it might instead be an address temporarily acquired by the mobile node (a co-located care-of address). A foreign agent, on the other hand, is a mobility agent that provides services to mobile nodes. See Sections 3.7 and 4.2.2 for additional details.

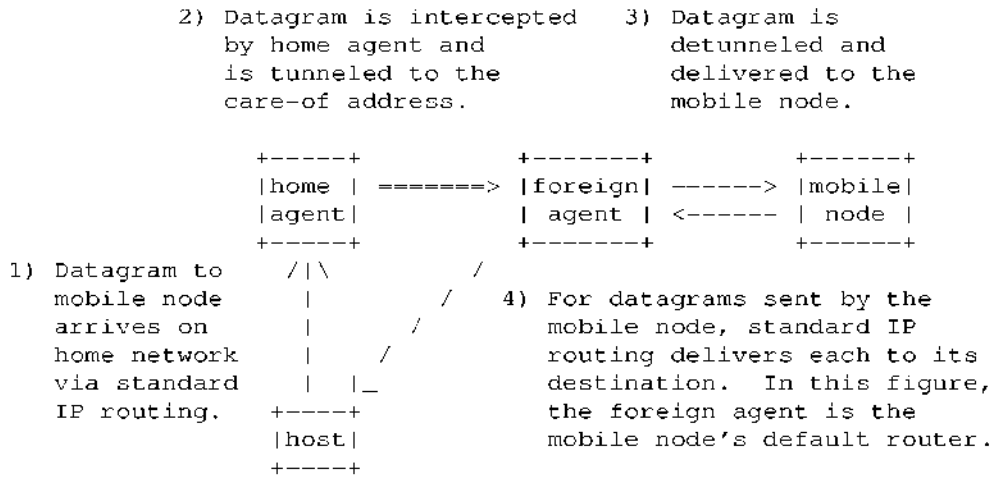
A home agent MUST be able to attract and intercept datagrams that are destined to the home address of any of its registered mobile nodes. Using the proxy and gratuitous ARP mechanisms described in Section 4.6, this requirement can be satisfied if the home agent has a network interface on the link indicated by the mobile node's home address. Other placements of the home agent relative to the mobile node's home location MAY also be possible using other mechanisms for intercepting datagrams destined to the mobile node's home address. Such placements are beyond the scope of this document.

Similarly, a mobile node and a prospective or current foreign agent MUST be able to exchange datagrams without relying on standard IP routing mechanisms; that is, those mechanisms which make forwarding decisions based upon the network-prefix of the destination address in the IP header. This requirement can be satisfied if the foreign agent and the visiting mobile node have an interface on the same link. In this case, the mobile node and foreign agent simply bypass their normal IP routing mechanism when sending datagrams to each other, addressing the underlying link-layer packets to their respective link-layer addresses. Other placements of the foreign agent relative to the mobile node MAY also be possible using other mechanisms to exchange datagrams between these nodes, but such placements are beyond the scope of this document.

If a mobile node is using a co-located care-of address (as described in (b) above), the mobile node MUST be located on the link identified by the network prefix of this care-of address. Otherwise, datagrams destined to the care-of address would be undeliverable.

For example, the figure below illustrates the routing of datagrams to and from a mobile node away from home, once the mobile node has registered with its home agent. In the figure below, the mobile node is using a foreign agent care-of address:

RED HAT 1010
RED HAT V. CAS



1.8. Specification Language

In this document, several words are used to signify the requirements of the specification. These words are often capitalized.

MUST — This word, or the adjective "required", means that the definition is an absolute requirement of the specification.

MUST NOT — This phrase means that the definition is an absolute prohibition of the specification.

SHOULD — This word, or the adjective "recommended", means that, in some circumstances, valid reasons may exist to ignore this item, but the full implications must be understood and carefully weighed before choosing a different course. Unexpected results may result otherwise.

MAY — This word, or the adjective "optional", means that this item is one of an allowed set of alternatives. An implementation which does not include this option MUST be prepared to interoperate with another implementation which does include the option.

RED HAT 1010
RED HAT V. CAS

silently discard
The implementation discards the datagram without further processing, and without indicating an error to the sender. The implementation SHOULD provide the capability of logging the error, including the contents of the discarded datagram, and SHOULD record the event in a statistics counter.

1.9. Message Format and Protocol Extensibility

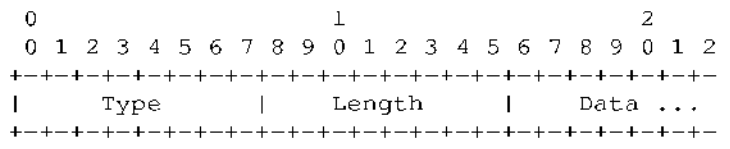
Mobile IP defines a set of new control messages, sent with UDP [17] using well-known port number 434. Currently, the following two message types are defined:

1 Registration Request
3 Registration Reply

Up-to-date values for the message types for Mobile IP control messages are specified in the most recent "Assigned Numbers" [20].

In addition, for Agent Discovery, Mobile IP makes use of the existing Router Advertisement and Router Solicitation messages defined for ICMP Router Discovery [4].

Mobile IP defines a general Extension mechanism to allow optional information to be carried by Mobile IP control messages or by ICMP Router Discovery messages. Each of these Extensions (with one exception) is encoded in the following Type-Length-Value format:



Type Indicates the particular type of Extension.

Length Indicates the length (in bytes) of the data field within this Extension. The length does NOT include the Type and Length bytes.

Data The particular data associated with this Extension. This field may be zero or more bytes in length. The format and length of the data field is determined by the type and length fields.

RED HAT 1010
RED HAT V. CAS

Extensions allow variable amounts of information to be carried within each datagram. The end of the list of Extensions is indicated by the total length of the IP datagram.

Two separately maintained sets of numbering spaces, from which Extension Type values are allocated, are used in Mobile IP:

- The first set consists of those Extensions which may appear only in Mobile IP control messages (those sent to and from UDP port number 434). Currently, the following Types are defined for Extensions appearing in Mobile IP control messages:

  32 Mobile-Home Authentication
  33 Mobile-Foreign Authentication
  34 Foreign-Home Authentication

- The second set consists of those extensions which may appear only in ICMP Router Discovery messages [4]. Currently, Mobile IP defines the following Types for Extensions appearing in ICMP Router Discovery messages:

  0 One-byte Padding (encoded with no Length nor Data field)
  16 Mobility Agent Advertisement
  19 Prefix-Lengths

Each individual Extension is described in detail in a separate section later in this document. Up-to-date values for these Extension Type numbers are specified in the most recent "Assigned Numbers" [20].

Due to the separation (orthogonality) of these sets, it is conceivable that two Extensions that are defined at a later date could have identical Type values, so long as one of the Extensions may be used only in Mobile IP control messages and the other may be used only in ICMP Router Discovery messages.

When an Extension numbered in either of these sets within the range 0 through 127 is encountered but not recognized, the message containing that Extension MUST be silently discarded. When an Extension numbered in the range 128 through 255 is encountered which is not recognized, that particular Extension is ignored, but the rest of the Extensions and message data MUST still be processed. The Length field of the Extension is used to skip the Data field in searching for the next Extension.

2. Agent Discovery

Agent Discovery is the method by which a mobile node determines whether it is currently connected to its home network or to a foreign network, and by which a mobile node can detect when it has moved from one network to another. When connected to a foreign network, the methods specified in this section also allow the mobile node to determine the foreign agent care-of address being offered by each foreign agent on that network.

Mobile IP extends ICMP Router Discovery [4] as its primary mechanism for Agent Discovery. An Agent Advertisement is formed by including a Mobility Agent Advertisement Extension in an ICMP Router Advertisement message (Section 2.1). An Agent Solicitation message is identical to an ICMP Router Solicitation, except that its IP TTL MUST be set to 1 (Section 2.2). This section describes the message formats and procedures by which mobile nodes, foreign agents, and home agents cooperate to realize Agent Discovery.

Agent Advertisement and Agent Solicitation may not be necessary for link layers that already provide this functionality. The method by which mobile nodes establish link-layer connections with prospective agents is outside the scope of this document (but see Appendix B). The procedures described below assume that such link-layer connectivity has already been established.

No authentication is required for Agent Advertisement and Agent Solicitation messages. They MAY be authenticated using the IP Authentication Header [1], which is unrelated to the messages described in this document. Further specification of the way in which Advertisement and Solicitation messages may be authenticated is outside of the scope of this document.

2.1. Agent Advertisement

Agent Advertisements are transmitted by a mobility agent to advertise its services on a link. Mobile nodes use these advertisements to determine their current point of attachment to the Internet. An Agent Advertisement is an ICMP Router Advertisement that has been extended to also carry an Mobility Agent Advertisement Extension (Section 2.1.1) and, optionally, a Prefix-Lengths Extension (Section 2.1.2), One-byte Padding Extension (Section 2.1.3), or other Extensions that might be defined in the future.

Within an Agent Advertisement message, ICMP Router Advertisement fields of the message are required to conform to the following additional specifications:

RED HAT 1010
RED HAT V. CAS

- Link-Layer Fields

  Destination Address
  The link-layer destination address of a unicast Agent Advertisement MUST be the same as the source link-layer address of the Agent Solicitation which prompted the Advertisement.

- IP Fields

  TTL The TTL for all Agent Advertisements MUST be set to 1.

  Destination Address
  As specified for ICMP Router Discovery [4], the IP destination address of an Agent Advertisement MUST be either the "all systems on this link" multicast address (224.0.0.1) [5] or the "limited broadcast" address (255.255.255.255). The subnet-directed broadcast address of the form <prefix>.<-1> cannot be used since mobile nodes will not generally know the prefix of the foreign network.

- ICMP Fields

  Code The Code field of the agent advertisement is interpreted as follows:

  0 The mobility agent handles common traffic -- that is, it acts as a router for IP datagrams not necessarily related to mobile nodes.
  16 The mobility agent does not route common traffic. However, all foreign agents MUST (minimally) forward to a default router any datagrams received from a registered mobile node (Section 4.2.2).

  Lifetime
  The maximum length of time that the Advertisement is considered valid in the absence of further Advertisements.

  Router Address(es)
  See Section 2.3.1 for a discussion of the addresses that may appear in this portion of the Agent Advertisement.

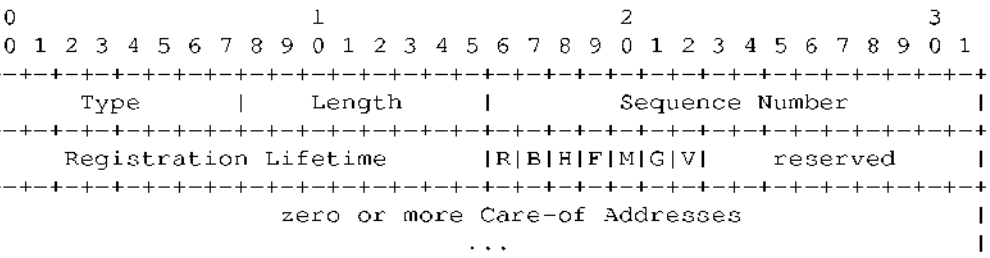
RED HAT 1010
RED HAT V. CAS

Num Addrs
The number of Router Addresses advertised in this message. Note that in an Agent Advertisement message, the number of router addresses specified in the ICMP Router Advertisement portion of the message MAY be set to 0. See Section 2.3.1 for details.

If sent periodically, the nominal interval at which Agent Advertisements are sent SHOULD be 1/3 of the advertisement Lifetime given in the ICMP header. This allows a mobile node to miss three successive advertisements before deleting the agent from its list of valid agents. The actual transmission time for each advertisement SHOULD be slightly randomized [4] in order to avoid synchronization and subsequent collisions with other Agent Advertisements that may be sent by other agents (or with other Router Advertisements sent by other routers). Note that this field has no relation to the "Registration Lifetime" field within the Mobility Agent Advertisement Extension defined below.

2.1.1. Mobility Agent Advertisement Extension

The Mobility Agent Advertisement Extension follows the ICMP Router Advertisement fields. It is used to indicate that an ICMP Router Advertisement message is also an Agent Advertisement being sent by a mobility agent. The Mobility Agent Advertisement Extension is defined as follows:



Type 16

Length (6 + 4*N), where N is the number of care-of addresses advertised.

Sequence Number
The count of Agent Advertisement messages sent since the agent was initialized (Section 2.3.2).

RED HAT 1010
RED HAT V. CAS

Registration Lifetime
: The longest lifetime (measured in seconds) that this agent is willing to accept in any Registration Request. A value of 0xffff indicates infinity. This field has no relation to the "Lifetime" field within the ICMP Router Advertisement portion of the Agent Advertisement.

R
: Registration required. Registration with this foreign agent (or another foreign agent on this link) is required rather than using a co-located care-of address.

B
: Busy. The foreign agent will not accept registrations from additional mobile nodes.

H
: Home agent. This agent offers service as a home agent on the link on which this Agent Advertisement message is sent.

F
: Foreign agent. This agent offers service as a foreign agent on the link on which this Agent Advertisement message is sent.

M
: Minimal encapsulation. This agent implements receiving tunneled datagrams that use minimal encapsulation [15].

G
: GRE encapsulation. This agent implements receiving tunneled datagrams that use GRE encapsulation [8].

V
: Van Jacobson header compression. This agent supports use of Van Jacobson header compression [10] over the link with any registered mobile node.

reserved
: Sent as zero; ignored on reception.

Care-of Address(es)
: The advertised foreign agent care-of address(es) provided by this foreign agent. An Agent Advertisement MUST include at least one care-of address if the 'F' bit is set. The number of care-of addresses present is determined by the Length field in the Extension.

A home agent MUST always be prepared to serve the mobile nodes for which it is the home agent. A foreign agent may at times be too busy to serve additional mobile nodes; even so, it must continue to send Agent Advertisements, so that any mobile nodes already registered with it will know that they have not moved out of range of the foreign agent and that the foreign agent has not failed. A foreign

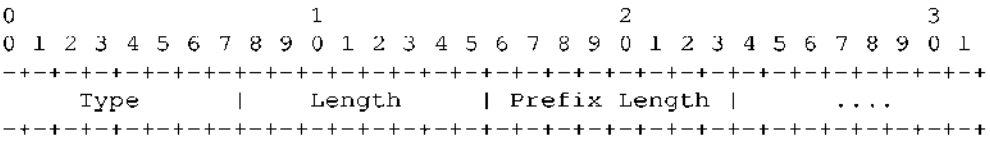
RED HAT 1010
RED HAT V. CAS

agent may indicate that it is "too busy" to allow new mobile nodes to register with it, by setting the 'B' bit in its Agent Advertisements. An Agent Advertisement message MUST NOT have the 'B' bit set if the 'F' bit is not also set, and at least one of the 'F' bit and the 'H' bit MUST be set in any Agent Advertisement message sent.

When a foreign agent wishes to require registration even from those mobile nodes which have acquired a co-located care-of address, it sets the 'R' bit to one. Because this bit applies only to foreign agents, an agent MUST NOT set the 'R' bit to one unless the 'F' bit is also set to one.

2.1.2. Prefix-Lengths Extension

The Prefix-Lengths Extension MAY follow the Mobility Agent Advertisement Extension. It is used to indicate the number of bits of network prefix that applies to each Router Address listed in the ICMP Router Advertisement portion of the Agent Advertisement. Note that the prefix lengths given DO NOT apply to care-of address(es) listed in the Mobility Agent Advertisement Extension. The Prefix-Lengths Extension is defined as follows:



Type 19 (Prefix-Lengths Extension)

Length N, where N is the value of the Num Addrs field in the ICMP Router Advertisement portion of the Agent Advertisement.

Prefix Length(s)
The number of leading bits that define the network number of the corresponding Router Address listed in the ICMP Router Advertisement portion of the message. The prefix length for each Router Address is encoded as a separate byte, in the order that the Router Addresses are listed in the ICMP Router Advertisement portion of the message.

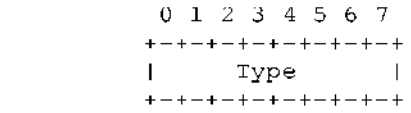
See Section 2.4.2 for information about how the Prefix Lengths Extension MAY be used by a mobile node when determining whether it has moved. See Appendix E for implementation details about the use of this Extension.

RED HAT 1010
RED HAT V. CAS

2.1.3. One-byte Padding Extension

Some IP protocol implementations insist upon padding ICMP messages to an even number of bytes. If the ICMP length of an Agent Advertisement is odd, this Extension MAY be included in order to make the ICMP length even. Note that this Extension is NOT intended to be a general-purpose Extension to be included in order to word- or long-align the various fields of the Agent Advertisement. An Agent Advertisement SHOULD NOT include more than one One-byte Padding Extension and if present, this Extension SHOULD be the last Extension in the Agent Advertisement.

Note that unlike other Extensions used in Mobile IP, the One-byte Padding Extension is encoded as a single byte, with no "Length" nor "Data" field present. The One-byte Padding Extension is defined as follows:



Type 0 (One-byte Padding Extension)

2.2. Agent Solicitation

An Agent Solicitation is identical to an ICMP Router Solicitation with the further restriction that the IP TTL Field MUST be set to 1.

2.3. Foreign Agent and Home Agent Considerations

Any mobility agent which cannot be discovered by a link-layer protocol MUST send Agent Advertisements. An agent which can be discovered by a link-layer protocol SHOULD also implement Agent Advertisements. However, the Advertisements need not be sent, except when the site policy requires registration with the agent (i.e., when the 'R' bit is set), or as a response to a specific Agent Solicitation. All mobility agents SHOULD respond to Agent Solicitations.

The same procedures, defaults, and constants are used in Agent Advertisement messages and Agent Solicitation messages as specified for ICMP Router Discovery [4], except that:

- a mobility agent MUST limit the rate at which it sends broadcast or multicast Agent Advertisements; a recommended maximum rate is once per second, AND

RED HAT 1010
RED HAT V. CAS

- a mobility agent that receives a Router Solicitation MUST NOT require that the IP Source Address is the address of a neighbor (i.e., an address that matches one of the router's own addresses on the arrival interface, under the subnet mask associated with that address of the router).

- a mobility agent MAY be configured to send Agent Advertisements only in response to an Agent Solicitation message.

If the home network is not a virtual network, then the home agent for any mobile node SHOULD be located on the link identified by the mobile node's home address, and Agent Advertisement messages sent by the home agent on this link MUST have the 'H' bit set. In this way, mobile nodes on their own home network will be able to determine that they are indeed at home. Any Agent Advertisement messages sent by the home agent on another link to which it may be attached (if it is a mobility agent serving more than one link), MUST NOT have the 'H' bit set, unless the home agent also serves as a home agent (to other mobile nodes) on that other link.

If the home network is a virtual network, the home network has no physical realization external to the home agent itself. In this case, there is no physical network link on which to send Agent Advertisement messages advertising the home agent. Mobile nodes for which this is the home network are always treated as being away from home.

On a particular subnet, either all mobility agents MUST include the Prefix-Lengths Extension or all of them MUST NOT include this Extension. Equivalently, it is prohibited for some agents on a given subnet to include the Extension but for others not to include it. Otherwise, one of the move detection algorithms designed for mobile nodes will not function properly (Section 2.4.2).

2.3.1. Advertised Router Addresses

The ICMP Router Advertisement portion of the Agent Advertisement MAY contain one or more router addresses. Thus, an agent MAY include one of its own addresses in the advertisement. A foreign agent MAY discourage use of this address as a default router by setting the preference to a low value and by including the address of another router in the advertisement (with a correspondingly higher preference). Nevertheless, a foreign agent MUST route datagrams it receives from registered mobile nodes (Section 4.2.2).

RED HAT 1010
352
RED HAT V. CAS

2.3.2. Sequence Numbers and Rollover Handling

The sequence number in Agent Advertisements ranges from 0 to 0xffff. After booting, an agent MUST use the number 0 for its first advertisement. Each subsequent advertisement MUST use the sequence number one greater, with the exception that the sequence number 0xffff MUST be followed by sequence number 256. In this way, mobile nodes can distinguish reductions in sequence numbers that result from reboots, from reductions that result in rollover of the sequence number after it attains the value 0xffff.

2.4. Mobile Node Considerations

Every mobile node MUST implement Agent Solicitation. Solicitations SHOULD only be sent in the absence of Agent Advertisements and when a care-of address has not been determined through a link-layer protocol or other means. The mobile node uses the same procedures, defaults, and constants for Agent Solicitation as specified for ICMP Router Solicitation messages [4], except that the mobile node MAY solicit more often than once every three seconds, and that a mobile node that is currently not connected to any foreign agent MAY solicit more times than MAX_SOLICITATIONS.

The rate at which a mobile node sends Solicitations MUST be limited by the mobile node. The mobile node MAY send three initial Solicitations at a maximum rate of one per second while searching for an agent. After this, the rate at which Solicitations are sent MUST be reduced so as to limit the overhead on the local link. Subsequent Solicitations MUST be sent using a binary exponential backoff mechanism, doubling the interval between consecutive Solicitations, up to a maximum interval. The maximum interval SHOULD be chosen appropriately based upon the characteristics of the media over which the mobile node is soliciting. This maximum interval SHOULD be at least one minute between Solicitations.

While still searching for an agent, the mobile node MUST NOT increase the rate at which it sends Solicitations unless it has received a positive indication that it has moved to a new link. After successfully registering with an agent, the mobile node SHOULD also increase the rate at which it will send Solicitations when it next begins searching for a new agent with which to register. The increased solicitation rate MAY revert to the maximum rate, but then MUST be limited in the manner described above. In all cases, the recommended solicitation intervals are nominal values. Mobile nodes MUST randomize their solicitation times around these nominal values as specified for ICMP Router Discovery [4].

RED HAT 1010
353
RED HAT V. CAS

Mobile nodes MUST process received Agent Advertisements. A mobile node can distinguish an Agent Advertisement message from other uses of the ICMP Router Advertisement message by examining the number of advertised addresses and the IP Total Length field. When the IP total length indicates that the ICMP message is longer than needed for the number of advertised addresses, the remaining data is interpreted as one or more Extensions. The presence of a Mobility Agent Advertisement Extension identifies the advertisement as an Agent Advertisement.

When multiple methods of agent discovery are in use, the mobile node SHOULD first attempt registration with agents including Mobility Agent Advertisement Extensions in their advertisements, in preference to those discovered by other means. This preference maximizes the likelihood that the registration will be recognized, thereby minimizing the number of registration attempts.

2.4.1. Registration Required

When the mobile node receives an Agent Advertisement with the 'R' bit set, the mobile node SHOULD register through the foreign agent, even when the mobile node might be able to acquire its own co-located care-of address. This feature is intended to allow sites to enforce visiting policies (such as accounting) which require exchanges of authorization.

2.4.2. Move Detection

Two primary mechanisms are provided for mobile nodes to detect when they have moved from one subnet to another. Other mechanisms MAY also be used. When the mobile node detects that it has moved, it SHOULD register (Section 3) with a suitable care-of address on the new foreign network. However, the mobile node MUST NOT register more frequently than once per second on average, as specified in Section 3.6.3.

RED HAT 1010
RED HAT V. CAS

2.4.2.1. Algorithm 1

The first method of move detection is based upon the Lifetime field within the main body of the ICMP Router Advertisement portion of the Agent Advertisement. A mobile node SHOULD record the Lifetime received in any Agent Advertisements, until that Lifetime expires. If the mobile node fails to receive another advertisement from the same agent within the specified Lifetime, it SHOULD assume that it has lost contact with that agent. If the mobile node has previously received an Agent Advertisement from another agent for which the Lifetime field has not yet expired, the mobile node MAY immediately attempt registration with that other agent. Otherwise, the mobile node SHOULD attempt to discover a new agent with which to register.

2.4.2.2. Algorithm 2

The second method uses network prefixes. The Prefix-Lengths Extension MAY be used in some cases by a mobile node to determine whether or not a newly received Agent Advertisement was received on the same subnet as the mobile node's current care-of address. If the prefixes differ, the mobile node MAY assume that it has moved. If a mobile node is currently using a foreign agent care-of address, the mobile node SHOULD NOT use this method of move detection unless both the current agent and the new agent include the Prefix-Lengths Extension in their respective Agent Advertisements; if this Extension is missing from one or both of the advertisements, this method of move detection SHOULD NOT be used. Similarly, if a mobile node is using a co-located care-of address, it SHOULD not use this method of move detection unless the new agent includes the Prefix-Lengths Extension in its Advertisement and the mobile node knows the network prefix of its current co-located care-of address. On the expiration of its current registration, if this method indicates that the mobile node has moved, rather than re-registering with its current care-of address, a mobile node MAY choose instead to register with a the foreign agent sending the new Advertisement with the different network prefix. The Agent Advertisement on which the new registration is based MUST NOT have expired according to its Lifetime field.

RED HAT 1010
RED HAT V. CAS

2.4.3. Returning Home

A mobile node can detect that it has returned to its home network when it receives an Agent Advertisement from its own home agent. If so, it SHOULD deregister with its home agent (Section 3). Before attempting to deregister, the mobile node SHOULD configure its routing table appropriately for its home network (Section 4.2.1). In addition, if the home network is using ARP [16], the mobile node MUST follow the procedures described in Section 4.6 with regard to ARP, proxy ARP, and gratuitous ARP.

2.4.4. Sequence Numbers and Rollover Handling

If a mobile node detects two successive values of the sequence number in the Agent Advertisements from the foreign agent with which it is registered, the second of which is less than the first and inside the range 0 to 255, the mobile node SHOULD register again. If the second value is less than the first but is greater than or equal to 256, the mobile node SHOULD assume that the sequence number has rolled over past its maximum value (0xffff), and that reregistration is not necessary (Section 2.3).

3. Registration

Mobile IP registration provides a flexible mechanism for mobile nodes to communicate their current reachability information to their home agent. It is the method by which mobile nodes:

- request forwarding services when visiting a foreign network,
- inform their home agent of their current care-of address,
- renew a registration which is due to expire, and/or
- deregister when they return home.

Registration messages exchange information between a mobile node, (optionally) a foreign agent, and the home agent. Registration creates or modifies a mobility binding at the home agent, associating the mobile node's home address with its care-of address for the specified Lifetime.

RED HAT 1010
RED HAT V. CAS

Several other (optional) capabilities are available through the registration procedure, which enable a mobile node to:

- maintain multiple simultaneous registrations, so that a copy of each datagram will be tunneled to each active care-of address
- deregister specific care-of addresses while retaining other mobility bindings, and
- discover the address of a home agent if the mobile node is not configured with this information.

3.1. Registration Overview

Mobile IP defines two different registration procedures, one via a foreign agent that relays the registration to the mobile node's home agent, and one directly with the mobile node's home agent. The following rules determine which of these two registration procedures to use in any particular circumstance:

- If a mobile node is registering a foreign agent care-of address, the mobile node MUST register via that foreign agent.
- If a mobile node is using a co-located care-of address, and receives an Agent Advertisement from a foreign agent on the link on which it is using this care-of address, the mobile node SHOULD register via that foreign agent (or via another foreign agent on this link) if the 'R' bit is set in the received Agent Advertisement message.
- If a mobile node is otherwise using a co-located care-of address, the mobile node MUST register directly with its home agent.
- If a mobile node has returned to its home network and is (de)registering with its home agent, the mobile node MUST register directly with its home agent.

Both registration procedures involve the exchange of Registration Request and Registration Reply messages (Sections 3.3 and 3.4). When registering via a foreign agent, the registration procedure requires the following four messages:

a) The mobile node sends a Registration Request to the prospective foreign agent to begin the registration process.

b) The foreign agent processes the Registration Request and then relays it to the home agent.

RED HAT 1010
RED HAT V. CAS

c) The home agent sends a Registration Reply to the foreign agent to grant or deny the Request.

d) The foreign agent processes the Registration Reply and then relays it to the mobile node to inform it of the disposition of its Request.

When the mobile node instead registers directly with its home agent, the registration procedure requires only the following two messages:

a) The mobile node sends a Registration Request to the home agent.

b) The home agent sends a Registration Reply to the mobile node, granting or denying the Request.

The registration messages defined in Sections 3.3 and 3.4 use the User Datagram Protocol (UDP) [17]. A nonzero UDP checksum SHOULD be included in the header, and MUST be checked by the recipient.

3.2. Authentication

Each mobile node, foreign agent, and home agent MUST be able to support a mobility security association for mobile entities, indexed by their SPI and IP address. In the case of the mobile node, this must be its Home Address. See Section 5.1 for requirements for support of authentication algorithms. Registration messages between a mobile node and its home agent MUST be authenticated with the Mobile-Home Authentication Extension (Section 3.5.2). This Extension immediately follows all non-authentication Extensions, except those foreign agent-specific Extensions which may be added to the message after the mobile node computes the authentication.

3.3. Registration Request

A mobile node registers with its home agent using a Registration Request message so that its home agent can create or modify a mobility binding for that mobile node (e.g., with a new lifetime). The Request may be relayed to the home agent by the foreign agent through which the mobile node is registering, or it may be sent directly to the home agent in the case in which the mobile node is registering a co-located care-of address.

IP fields:

Source Address Typically the interface address from which the message is sent.

RED HAT 1010
RED HAT V. CAS

Destination Address Typically that of the foreign agent or the home agent.

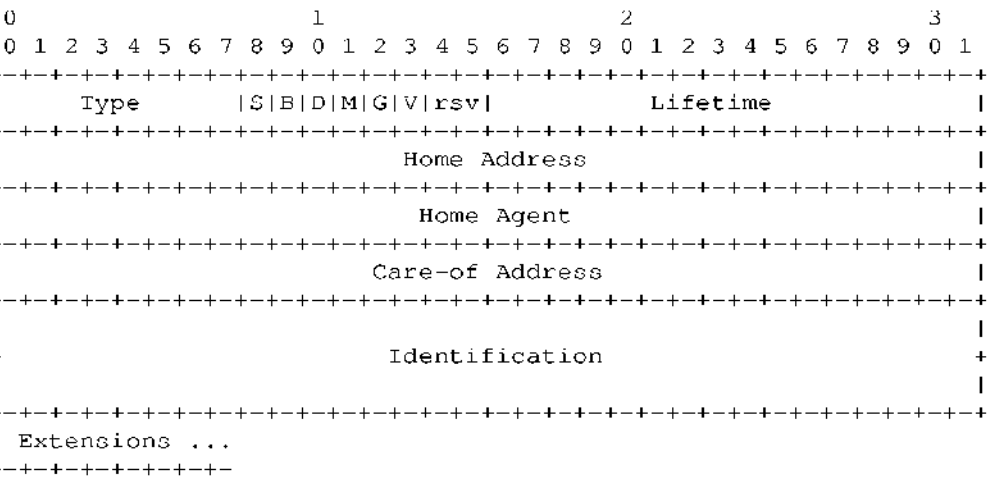
See Sections 3.6.1.1 and 3.7.2.2 for details.

UDP fields:

Source Port variable

Destination Port 434

The UDP header is followed by the Mobile IP fields shown below:



Type 1 (Registration Request)

S Simultaneous bindings. If the 'S' bit is set, the mobile node is requesting that the home agent retain its prior mobility bindings, as described in Section 3.6.1.2.

B Broadcast datagrams. If the 'B' bit is set, the mobile node requests that the home agent tunnel to it any broadcast datagrams that it receives on the home network, as described in Section 4.3.

D Decapsulation by mobile node. If the 'D' bit is set, the mobile node will itself decapsulate datagrams which are sent to the care-of address. That is, the mobile node is using a co-located care-of address.

RED HAT 1010
RED HAT V. CAS

M Minimal encapsulation. If the 'M' bit is set, the mobile node requests that its home agent use minimal encapsulation [15] for datagrams tunneled to the mobile node.

G GRE encapsulation. If the 'G' bit is set, the mobile node requests that its home agent use GRE encapsulation [8] for datagrams tunneled to the mobile node.

V The mobile node requests that its mobility agent use Van Jacobson header compression [10] over its link with the mobile node.

rsv Reserved bits; sent as zero

Lifetime
The number of seconds remaining before the registration is considered expired. A value of zero indicates a request for deregistration. A value of 0xffff indicates infinity.

Home Address
The IP address of the mobile node.

Home Agent
The IP address of the mobile node's home agent.

Care-of Address
The IP address for the end of the tunnel.

Identification
A 64-bit number, constructed by the mobile node, used for matching Registration Requests with Registration Replies, and for protecting against replay attacks of registration messages. See Sections 5.4 and 5.6.

Extensions
The fixed portion of the Registration Request is followed by one or more of the Extensions listed in Section 3.5. The Mobile-Home Authentication Extension MUST be included in all Registration Requests. See Sections 3.6.1.3 and 3.7.2.2 for information on the relative order in which different extensions, when present, MUST be placed in a Registration Request message.

RED HAT 1010
RED HAT V. CAS

3.4. Registration Reply

A mobility agent returns a Registration Reply message to a mobile node which has sent a Registration Request (Section 3.3) message. If the mobile node is requesting service from a foreign agent, that foreign agent will receive the Reply from the home agent and subsequently relay it to the mobile node. The Reply message contains the necessary codes to inform the mobile node about the status of its Request, along with the lifetime granted by the home agent, which MAY be smaller than the original Request.

The foreign agent MUST NOT increase the Lifetime selected by the mobile node in the Registration Request, since the Lifetime is covered by the Mobile-Home Authentication Extension, which cannot be correctly (re)computed by the foreign agent. The home agent MUST NOT increase the Lifetime selected by the mobile node in the Registration Request, since doing so could increase it beyond the maximum Registration Lifetime allowed by the foreign agent. If the Lifetime received in the Registration Reply is greater than that in the Registration Request, the Lifetime in the Request MUST be used. When the Lifetime received in the Registration Reply is less than that in the Registration Request, the Lifetime in the Reply MUST be used.

IP fields:

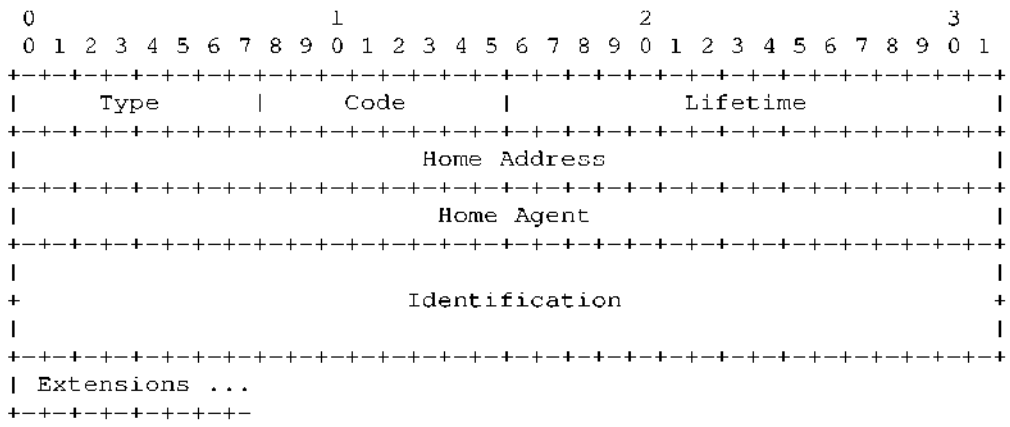
Source Address	Typically copied from the destination address of the Registration Request to which the agent is replying. See Sections 3.7.2.3 and 3.8.3.1 for complete details.
Destination Address	Copied from the source address of the Registration Request to which the agent is replying

UDP fields:

Source Port	<variable>
Destination Port	Copied from the source port of the corresponding Registration Request (Section 3.7.1).

RED HAT 1010
RED HAT V. CAS

The UDP header is followed by the Mobile IP fields shown below:



Type 3 (Registration Reply)

Code A value indicating the result of the Registration Request. See below for a list of currently defined Code values.

Lifetime
If the Code field indicates that the registration was accepted, the Lifetime field is set to the number of seconds remaining before the registration is considered expired. A value of zero indicates that the mobile node has been deregistered. A value of 0xffff indicates infinity. If the Code field indicates that the registration was denied, the contents of the Lifetime field are unspecified and MUST be ignored on reception.

Home Address
The IP address of the mobile node.

Home Agent
The IP address of the mobile node's home agent.

RED HAT 1010
RED HAT V. CAS

Identification
: A 64-bit number used for matching Registration Requests with Registration Replies, and for protecting against replay attacks of registration messages. The value is based on the Identification field from the Registration Request message from the mobile node, and on the style of replay protection used in the security context between the mobile node and its home agent (defined by the mobility security association between them, and SPI value in the Mobile-Home Authentication Extension). See Sections 5.4 and 5.6.

Extensions
: The fixed portion of the Registration Reply is followed by one or more of the Extensions listed in Section 3.5. The Mobile-Home Authentication Extension MUST be included in all Registration Replies returned by the home agent. See Sections 3.7.2.2 and 3.8.3.3 for rules on placement of extensions to Reply messages.

The following values are defined for use within the Code field.
Registration successful:

0 registration accepted
1 registration accepted, but simultaneous mobility bindings unsupported

Registration denied by the foreign agent:

64 reason unspecified
65 administratively prohibited
66 insufficient resources
67 mobile node failed authentication
68 home agent failed authentication
69 requested Lifetime too long
70 poorly formed Request
71 poorly formed Reply
72 requested encapsulation unavailable
73 requested Van Jacobson compression unavailable
80 home network unreachable (ICMP error received)
81 home agent host unreachable (ICMP error received)
82 home agent port unreachable (ICMP error received)
88 home agent unreachable (other ICMP error received)

Registration denied by the home agent:

128 reason unspecified
129 administratively prohibited
130 insufficient resources
131 mobile node failed authentication
132 foreign agent failed authentication
133 registration Identification mismatch
134 poorly formed Request
135 too many simultaneous mobility bindings
136 unknown home agent address

Up-to-date values of the Code field are specified in the most recent "Assigned Numbers" [20].

3.5. Registration Extensions

3.5.1. Computing Authentication Extension Values

The Authenticator value computed for each authentication Extension MUST protect the following fields from the registration message:

- the UDP payload (that is, the Registration Request or Registration Reply data),
- all prior Extensions in their entirety, and
- the Type and Length of this Extension.

The default authentication algorithm uses keyed-MD5 [21] in "prefix+suffix" mode to compute a 128-bit "message digest" of the registration message. The default authenticator is a 128-bit value computed as the MD5 checksum over the the following stream of bytes:

- the shared secret defined by the mobility security association between the nodes and by SPI value specified in the authentication Extension, followed by
- the protected fields from the registration message, in the order specified above, followed by
- the shared secret again.

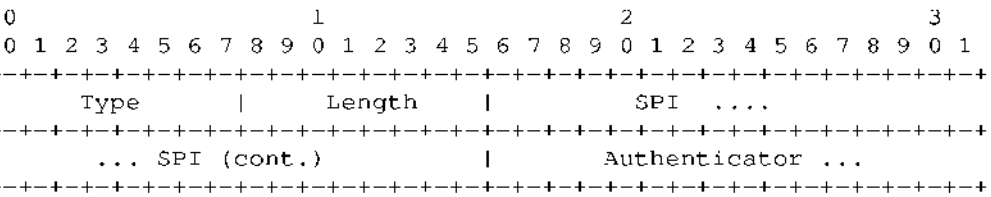
Note that the Authenticator field itself and the UDP header are NOT included in the computation of the default Authenticator value. See Section 5.1 for information about support requirements for message authentication codes, which are to be used with the various authentication Extensions.

RED HAT 1010
RED HAT V. CAS

The Security Parameter Index (SPI) within any of the authentication Extensions defines the security context which is used to compute the Authenticator value and which MUST be used by the receiver to check that value. In particular, the SPI selects the authentication algorithm and mode (Section 5.1) and secret (a shared key, or appropriate public/private key pair) used in computing the Authenticator. In order to ensure interoperability between different implementations of the Mobile IP protocol, an implementation MUST be able to associate any SPI value with any authentication algorithm and mode which it implements. In addition, all implementations of Mobile IP MUST implement the default authentication algorithm (keyed-MD5) and mode ("prefix+suffix") defined above.

3.5.2. Mobile-Home Authentication Extension

Exactly one Mobile-Home Authentication Extension MUST be present in all Registration Requests and Registration Replies, and is intended to eliminate problems [2] which result from the uncontrolled propagation of remote redirects in the Internet. The location of the extension marks the end of the authenticated data.

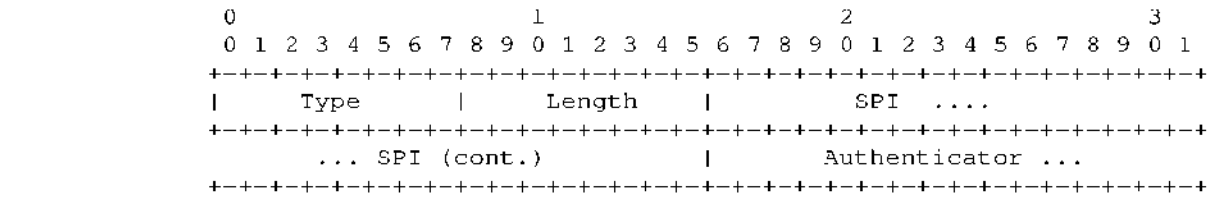


Type	32
Length	4 plus the number of bytes in the Authenticator.
SPI	Security Parameter Index (4 bytes). An opaque identifier (see Section 1.6).
Authenticator	(variable length) (See Section 3.5.1.)

3.5.3. Mobile-Foreign Authentication Extension

This Extension MAY be included in Registration Requests and Replies in cases in which a mobility security association exists between the mobile node and the foreign agent. See Section 5.1 for information about support requirements for message authentication codes.

RED HAT 1010
RED HAT V. CAS



Type 33

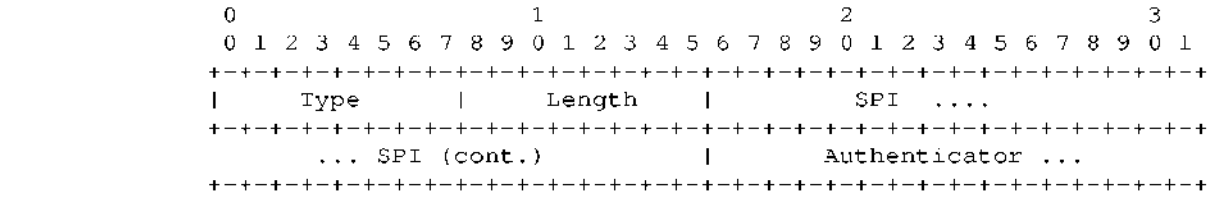
Length 4 plus the number of bytes in the Authenticator.

SPI Security Parameter Index (4 bytes). An opaque identifier (see Section 1.6).

Authenticator (variable length) (See Section 3.5.1.)

3.5.4. Foreign-Home Authentication Extension

This Extension MAY be included in Registration Requests and Replies in cases in which a mobility security association exists between the foreign agent and the home agent. See Section 5.1 for information about support requirements for message authentication codes.



Type 34

Length 4 plus the number of bytes in the Authenticator.

SPI Security Parameter Index (4 bytes). An opaque identifier (see Section 1.6).

Authenticator (variable length) (See Section 3.5.1.)

3.6. Mobile Node Considerations

A mobile node MUST be configured with its home address, a netmask, and a mobility security association for each home agent. In addition, a mobile node MAY be configured with the IP address of one or more of its home agents; otherwise, the mobile node MAY discover a home agent using the procedures described in Section 3.6.1.2.

RED HAT 1010
RED HAT V. CAS

For each pending registration, the mobile node maintains the following information:

- the link-layer address of the foreign agent to which the Registration Request was sent, if applicable,
- the IP destination address of the Registration Request,
- the care-of address used in the registration,
- the Identification value sent in the registration,
- the originally requested Lifetime, and
- the remaining Lifetime of the pending registration.

A mobile node SHOULD initiate a registration whenever it detects a change in its network connectivity. See Section 2.4.2 for methods by which mobile nodes MAY make such a determination. When it is away from home, the mobile node's Registration Request allows its home agent to create or modify a mobility binding for it. When it is at home, the mobile node's (de)Registration Request allows its home agent to delete any previous mobility binding(s) for it. A mobile node operates without the support of mobility functions when it is at home.

There are other conditions under which the mobile node SHOULD (re)register with its foreign agent, such as when the mobile node detects that the foreign agent has rebooted (as specified in Section 2.4.4) and when the current registration's Lifetime is near expiration.

In the absence of link-layer indications of changes in point of attachment, Agent Advertisements from new agents SHOULD NOT cause a mobile node to attempt a new registration, if its current registration has not expired and it is still also receiving Agent Advertisements from the foreign agent with which it is currently registered. In the absence of link-layer indications, a mobile node MUST NOT attempt to register more often than once per second.

A mobile node MAY register with a different agent when transport-layer protocols indicate excessive retransmissions. A mobile node MUST NOT consider reception of an ICMP Redirect from a foreign agent that is currently providing service to it as reason to register with a new foreign agent. Within these constraints, the mobile node MAY register again at any time.

Appendix D shows some examples of how the fields in registration messages would be set up in some typical registration scenarios.

RED HAT 1010
RED HAT V. CAS

3.6.1. Sending Registration Requests

The following sections specify details for the values the mobile node MUST supply in the fields of Registration Request messages.

3.6.1.1. IP Fields

This section provides the specific rules by which mobile nodes pick values for the IP header fields of a Registration Request.

IP Source Address:

- When registering on a foreign network with a co-located care-of address, the IP source address MUST be the care-of address.

- In all other circumstances, the IP source address MUST be the mobile node's home address.

IP Destination Address:

- When the mobile node has discovered the agent with which it is registering, through some means (e.g., link-layer) that does not provide the IP address of the agent (the IP address of the agent is unknown to the mobile node), then the "All Mobility Agents" multicast address (224.0.0.11) MUST be used. In this case, the mobile node MUST use the agent's link-layer unicast address in order to deliver the datagram to the correct agent.

- When registering with a foreign agent, the address of the agent as learned from the IP source address of the corresponding Agent Advertisement MUST be used. In addition, when transmitting this Registration Request message, the mobile node MUST use a link-layer destination address copied from the link-layer source address of the Agent Advertisement message in which it learned this foreign agent's IP address.

- When the mobile node is registering directly with its home agent and knows the (unicast) IP address of its home agent, the destination address MUST be set to this address.

RED HAT 1010
RED HAT V. CAS

- If the mobile node is registering directly with its home agent, but does not know the IP address of its home agent, the mobile node may use dynamic home agent address resolution to automatically determine the IP address of its home agent (Section 3.6.1.2). In this case, the IP destination address is set to the subnet-directed broadcast address of the mobile node's home network. This address MUST NOT be used as the destination IP address if the mobile node is registering via a foreign agent, although it MAY be used as the Home Agent address in the body of the Registration Request when registering via a foreign agent.

IP Time to Live:

- The IP TTL field MUST be set to 1 if the IP destination address is set to the "All Mobility Agents" multicast address as described above. Otherwise a suitable value should be chosen in accordance with standard IP practice [19].

3.6.1.2. Registration Request Fields

This section provides specific rules by which mobile nodes pick values for the fields within the fixed portion of a Registration Request.

A mobile node MAY set the 'S' bit in order to request that the home agent maintain prior mobility binding(s). Otherwise, the home agent deletes any previous binding(s) and replaces them with the new binding specified in the Registration Request. Multiple simultaneous mobility bindings are likely to be useful when a mobile node using at least one wireless network interface moves within wireless transmission range of more than one foreign agent. IP explicitly allows duplication of datagrams. When the home agent allows simultaneous bindings, it will tunnel a separate copy of each arriving datagram to each care-of address, and the mobile node will receive multiple copies of datagrams destined to it.

The mobile node SHOULD set the 'D' bit if it is registering with a co-located care-of address. Otherwise, the 'D' bit MUST NOT be set.

A mobile node MAY set the 'B' bit to request its home agent to forward to it, a copy of broadcast datagrams received by its home agent from the home network. The method used by the home agent to forward broadcast datagrams depends on the type of care-of address registered by the mobile node, as determined by the 'D' bit in the mobile node's Registration Request:

RED HAT 1010
RED HAT V. CAS

- If the 'D' bit is set, then the mobile node has indicated that it will decapsulate any datagrams tunneled to this care-of address itself (the mobile node is using a co-located care-of address). In this case, to forward such a received broadcast datagram to the mobile node, the home agent MUST tunnel it to this care-of address. The mobile node de-tunnels the received datagram in the same way as any other datagram tunneled directly to it.

- If the 'D' bit is NOT set, then the mobile node has indicated that it is using a foreign agent care-of address, and that the foreign agent will thus decapsulate arriving datagrams before forwarding them to the mobile node. In this case, to forward such a received broadcast datagram to the mobile node, the home agent MUST first encapsulate the broadcast datagram in a unicast datagram addressed to the mobile node's home address, and then MUST tunnel this resulting datagram to the mobile node's care-of address.

  When decapsulated by the foreign agent, the inner datagram will thus be a unicast IP datagram addressed to the mobile node, identifying to the foreign agent the intended destination of the encapsulated broadcast datagram, and will be delivered to the mobile node in the same way as any tunneled datagram arriving for the mobile node. The foreign agent MUST NOT decapsulate the encapsulated broadcast datagram and MUST NOT use a local network broadcast to transmit it to the mobile node. The mobile node thus MUST decapsulate the encapsulated broadcast datagram itself, and thus MUST NOT set the 'B' bit in its Registration Request in this case unless it is capable of decapsulating datagrams.

The mobile node MAY request alternative forms of encapsulation by setting the 'M' bit and/or the 'G' bit, but only if the mobile node is decapsulating its own datagrams (the mobile node is using a co-located care-of address) or if its foreign agent has indicated support for these forms of encapsulation by setting the corresponding bits in the Mobility Agent Advertisement Extension of an Agent Advertisement received by the mobile node. Otherwise, the mobile node MUST NOT set these bits.

The Lifetime field is chosen as follows:

- If the mobile node is registering with a foreign agent, the Lifetime SHOULD NOT exceed the value in the Registration Lifetime field of the Agent Advertisement message received from the foreign agent. When the method by which the care-of address is learned does not include a Lifetime, the default ICMP Router Advertisement Lifetime (1800 seconds) MAY be used.

RED HAT 1010
RED HAT V. CAS

- The mobile node MAY ask a home agent to delete a particular mobility binding, by sending a Registration Request with the care-of address for this binding, with the Lifetime field set to zero (Section 3.8.2).

- Similarly, a Lifetime of zero is used when the mobile node deregisters all care-of addresses, such as upon returning home.

The Home Agent field MUST be set to the address of the mobile node's home agent, if the mobile node knows this address. Otherwise, the mobile node MAY use dynamic home agent address resolution to learn the address of its home agent. In this case, the mobile node MUST set the Home Agent field to the subnet-directed broadcast address of the mobile node's home network. Each home agent receiving such a Registration Request with a broadcast destination address MUST reject the mobile node's registration and SHOULD return a rejection Registration Reply indicating its unicast IP address for use by the mobile node in a future registration attempt.

The Care-of Address field MUST be set to the value of the particular care-of address that the mobile node wishes to (de)register. In the special case in which a mobile node wishes to deregister all care-of addresses, it MUST set this field to its home address.

The mobile node chooses the Identification field in accordance with the style of replay protection it uses with its home agent. This is part of the mobility security association the mobile node shares with its home agent. See Section 5.6 for the method by which the mobile node computes the Identification field.

3.6.1.3. Extensions

This section describes the ordering of any mandatory and any optional Extensions that a mobile node appends to a Registration Request. This following ordering MUST be followed:

a) The IP header, followed by the UDP header, followed by the fixed-length portion of the Registration Request, followed by

b) If present, any non-authentication Extensions expected to be used by the home agent (which may or may not also be used by the foreign agent), followed by

c) The Mobile-Home Authentication Extension, followed by

d) If present, any non-authentication Extensions used only by the foreign agent, followed by

RED HAT 1010
RED HAT V. CAS

e) The Mobile-Foreign Authentication Extension, if present.

Note that items (a) and (c) MUST appear in every Registration Request sent by the mobile node. Items (b), (d), and (e) are optional. However, item (e) MUST be included when the mobile node and the foreign agent share a mobility security association.

3.6.2. Receiving Registration Replies

Registration Replies will be received by the mobile node in response to its Registration Requests. Registration Replies generally fall into three categories:

- the registration was accepted,
- the registration was denied by the foreign agent, or
- the registration was denied by the home agent.

The remainder of this section describes the Registration Reply handling by a mobile node in each of these three categories.

3.6.2.1. Validity Checks

Registration Replies with an invalid, non-zero UDP checksum MUST be silently discarded.

In addition, the low-order 32 bits of the Identification field in the Registration Reply MUST be compared to the low-order 32 bits of the Identification field in the most recent Registration Request sent to the replying agent. If they do not match, the Reply MUST be silently discarded.

Also, the authentication in the Registration Reply MUST be checked. That is, the mobile node MUST check for the presence of a valid authentication Extension, acting in accordance with the Code field in the Reply. The rules are as follows:

a) If the mobile node and the foreign agent share a mobility security association, exactly one Mobile-Foreign Authentication Extension MUST be present in the Registration Reply, and the mobile node MUST check the Authenticator value in the Extension. If no Mobile-Foreign Authentication Extension is found, or if more than one Mobile-Foreign Authentication Extension is found, or if the Authenticator is invalid, the mobile node MUST silently discard the Reply and SHOULD log the event as a security exception.

RED HAT 1010
RED HAT V. CAS

b) If the Code field indicates that service is denied by the home agent, or if the Code field indicates that the registration was accepted by the home agent, exactly one Mobile-Home Authentication Extension MUST be present in the Registration Reply, and the mobile node MUST check the Authenticator value in the Extension. If no Mobile-Home Authentication Extension is found, or if more than one Mobile-Home Authentication Extension is found, or if the Authenticator is invalid, the mobile node MUST silently discard the Reply and SHOULD log the event as a security exception.

If the Code field indicates an authentication failure, either at the foreign agent or the home agent, then it is quite possible that any authenticators in the Registration Reply will also be in error. This could happen, for example, if the shared secret between the mobile node and home agent was erroneously configured. The mobile node SHOULD log such errors as security exceptions.

3.6.2.2. Registration Request Accepted

If the Code field indicates that the request has been accepted, the mobile node SHOULD configure its routing table appropriately for its current point of attachment (Section 4.2.1).

If the mobile node is returning to its home network and that network is one which implements ARP, the mobile node MUST follow the procedures described in Section 4.6 with regard to ARP, proxy ARP, and gratuitous ARP.

If the mobile node has registered on a foreign network, it SHOULD re-register before the expiration of the Lifetime of its registration. As described in Section 3.6, for each pending Registration Request, the mobile node MUST maintain the remaining lifetime of this pending registration, as well as the original Lifetime from the Registration Request. When the mobile node receives a valid Registration Reply, the mobile node MUST decrease its view of the remaining lifetime of the registration by the amount by which the home agent decreased the originally requested Lifetime. This procedure is equivalent to the mobile node starting a timer for the granted Lifetime at the time it sent the Registration Request, even though the granted Lifetime is not known to the mobile node until the Registration Reply is received. Since the Registration Request is certainly sent before the home agent begins timing the registration Lifetime (also based on the granted Lifetime), this procedure ensures that the mobile node will re-register before the home agent expires and deletes the registration, in spite of possibly non-negligible transmission delays for the original Registration

RED HAT 1010
RED HAT V. CAS

Request and Reply that started the timing of the Lifetime at the mobile node and its home agent.

3.6.2.3. Registration Request Denied

If the Code field indicates that service is being denied, the mobile node SHOULD log the error. In certain cases the mobile node may be able to "repair" the error. These include:

Code 69: (Denied by foreign agent, Lifetime too long)

In this case, the Lifetime field in the Registration Reply will contain the maximum Lifetime value which that foreign agent is willing to accept in any Registration Request. The mobile node MAY attempt to register with this same agent, using a Lifetime in the Registration Request that MUST be less than or equal to the value specified in the Reply.

Code 133: (Denied by home agent, Identification mismatch)

In this case, the Identification field in the Registration Reply will contain a value that allows the mobile node to synchronize with the home agent, based upon the style of replay protection in effect (Section 5.6). The mobile node MUST adjust the parameters it uses to compute the Identification field based upon the information in the Registration Reply, before issuing any future Registration Requests.

Code 136: (Denied by home agent, Unknown home agent address)

This code is returned by a home agent when the mobile node is performing dynamic home agent address resolution as described in Sections 3.6.1.1 and 3.6.1.2. In this case, the Home Agent field within the Reply will contain the unicast IP address of the home agent returning the Reply. The mobile node MAY then attempt to register with this home agent in future Registration Requests. In addition, the mobile node SHOULD adjust the parameters it uses to compute the Identification field based upon the corresponding field in the Registration Reply, before issuing any future Registration Requests.

3.6.3. Registration Retransmission

When no Registration Reply has been received within a reasonable time, another Registration Request MAY be transmitted. When timestamps are used, a new registration Identification is chosen for each retransmission; thus it counts as a new registration. When nonces are used, the unanswered Request is retransmitted unchanged;

RED HAT 1010
RED HAT V. CAS

thus the retransmission does not count as a new registration (Section 5.6). In this way a retransmission will not require the home agent to resynchronize with the mobile node by issuing another nonce in the case in which the original Registration Request (rather than its Registration Reply) was lost by the network.

The maximum time until a new Registration Request is sent SHOULD be no greater than the requested Lifetime of the Registration Request. The minimum value SHOULD be large enough to account for the size of the messages, twice the round trip time for transmission to the home agent, and at least an additional 100 milliseconds to allow for processing the messages before responding. The round trip time for transmission to the home agent will be at least as large as the time required to transmit the messages at the link speed of the mobile node's current point of attachment. Some circuits add another 200 milliseconds of satellite delay in the total round trip time to the home agent. The minimum time between Registration Requests MUST NOT be less than 1 second. Each successive retransmission timeout period SHOULD be at least twice the previous period, as long as that is less than the maximum as specified above.

3.7. Foreign Agent Considerations

The foreign agent plays a mostly passive role in Mobile IP registration. It relays Registration Requests between mobile nodes and home agents, and, when it provides the care-of address, decapsulates datagrams for delivery to the mobile node. It SHOULD also send periodic Agent Advertisement messages to advertise its presence as described in Section 2.3, if not detectable by link-layer means.

A foreign agent MUST NOT transmit a Registration Request except when relaying a Registration Request received from a mobile node, to the mobile node's home agent. A foreign agent MUST NOT transmit a Registration Reply except when relaying a Registration Reply received from a mobile node's home agent, or when replying to a Registration Request received from a mobile node in the case in which the foreign agent is denying service to the mobile node. In particular, a foreign agent MUST NOT generate a Registration Request or Reply because a mobile node's registration Lifetime has expired. A foreign agent also MUST NOT originate a Registration Request message that asks for deregistration of a mobile node; however, it MUST relay valid (de)Registration Requests originated by a mobile node.

RED HAT 1010
RED HAT V. CAS

3.7.1. Configuration and Registration Tables

Each foreign agent MUST be configured with a care-of address. In addition, for each pending or current registration, the foreign agent MUST maintain a visitor list entry containing the following information obtained from the mobile node's Registration Request:

- the link-layer source address of the mobile node
- the IP Source Address (the mobile node's Home Address)
- the IP Destination Address (as specified in 3.6.2.3)
- the UDP Source Port
- the Home Agent address
- the Identification field
- the requested registration Lifetime, and
- the remaining Lifetime of the pending or current registration.

As with any node on the Internet, a foreign agent MAY also share mobility security associations with any other nodes. When relaying a Registration Request from a mobile node to its home agent, if the foreign agent shares a mobility security association with the home agent, it MUST add a Foreign-Home Authentication Extension to the Request and MUST check the required Foreign-Home Authentication Extension in the Registration Reply from the home agent (Sections 3.3 and 3.4). Similarly, when receiving a Registration Request from a mobile node, if the foreign agent shares a mobility security association with the mobile node, it MUST check the required Mobile-Foreign Authentication Extension in the Request and MUST add a Mobile-Foreign Authentication Extension to the Registration Reply to the mobile node.

3.7.2. Receiving Registration Requests

If the foreign agent accepts a Registration Request from a mobile node, it then MUST relay the Request to the indicated home agent. Otherwise, if the foreign agent denies the Request, it MUST send a Registration Reply to the mobile node with an appropriate denial Code, except in cases where the foreign agent would be required to send out more than one such denial per second to the same mobile node. The following sections describe this behavior in more detail.

If a foreign agent receives a Registration Request from a mobile node in its visitor list, the existing visitor list entry for the mobile node SHOULD NOT be deleted or modified until the foreign agent receives a valid Registration Reply from the home agent with a Code indicating success. The foreign agent MUST record the new pending

RED HAT 1010
RED HAT V. CAS

Request separately from the existing visitor list entry for the mobile node. If the Registration Request requests deregistration, the existing visitor list entry for the mobile node SHOULD NOT be deleted until the foreign agent has received a successful Registration Reply. If the Registration Reply indicates that the Request (for registration or deregistration) was denied by the home agent, the existing visitor list entry for the mobile node MUST NOT be modified as a result of receiving the Registration Reply.

3.7.2.1. Validity Checks

Registration Requests with an invalid, non-zero UDP checksum MUST be silently discarded.

Also, the authentication in the Registration Request MUST be checked. If the foreign agent and the mobile node share a mobility security association, exactly one Mobile-Foreign Authentication Extension MUST be present in the Registration Request, and the foreign agent MUST check the Authenticator value in the Extension. If no Mobile-Foreign Authentication Extension is found, or if more than one Mobile-Foreign Authentication Extension is found, or if the Authenticator is invalid, the foreign agent MUST silently discard the Request and SHOULD log the event as a security exception. The foreign agent also SHOULD send a Registration Reply to the mobile node with Code 67.

3.7.2.2. Forwarding a Valid Request to the Home Agent

If the foreign agent accepts the mobile node's Registration Request, it MUST relay the Request to the mobile node's home agent as specified in the Home Agent field of the Registration Request. The foreign agent MUST NOT modify any of the fields beginning with the fixed portion of the Registration Request up through and including the Mobile-Home Authentication Extension. Otherwise, an authentication failure is very likely to occur at the home agent. In addition, the foreign agent proceeds as follows:

- It MUST process and remove any Extensions following the Mobile-Home Authentication Extension,
- It MAY append any of its own non-authentication Extensions of relevance to the home agent, if applicable, and
- It MUST append the Foreign-Home Authentication Extension, if the foreign agent shares a mobility security association with the home agent.

RED HAT 1010
RED HAT V. CAS

Specific fields within the IP header and the UDP header of the relayed Registration Request MUST be set as follows:

IP Source Address
: The foreign agent's address on the interface from which the message will be sent.

IP Destination Address
: Copied from the Home Agent field within the Registration Request.

UDP Source Port
: <variable>

UDP Destination Port
: 434

After forwarding a valid Registration Request to the home agent, the foreign agent MUST begin timing the remaining lifetime of the pending registration based on the Lifetime in the Registration Request. If this lifetime expires before receiving a valid Registration Reply, the foreign agent MUST delete its visitor list entry for this pending registration.

3.7.2.3. Denying Invalid Requests

If the foreign agent denies the mobile node's Registration Request for any reason, it SHOULD send the mobile node a Registration Reply with a suitable denial Code. In such a case, the Home Address, Home Agent, and Identification fields within the Registration Reply are copied from the corresponding fields of the Registration Request.

If the Reserved field is nonzero, the foreign agent MUST deny the Request and SHOULD return a Registration Reply with status code 70 to the mobile node. If the Request is being denied because the requested Lifetime is too long, the foreign agent sets the Lifetime in the Reply to the maximum Lifetime value it is willing to accept in any Registration Request, and sets the Code field to 69. Otherwise, the Lifetime SHOULD be copied from the Lifetime field in the Request.

Specific fields within the IP header and the UDP header of the Registration Reply MUST be set as follows:

IP Source Address
: Copied from the IP Destination Address of Registration Request, unless the "All Agents Multicast" address was used. In this case, the foreign agent's address (on the interface from which the message will be sent) MUST be

RED HAT 1010
RED HAT V. CAS

used.

IP Destination Address
Copied from the IP Source Address of the Registration Request.

UDP Source Port
434

UDP Destination Port
Copied from the UDP Source Port of the Registration Request.

3.7.3. Receiving Registration Replies

The foreign agent updates its visitor list when it receives a valid Registration Reply from a home agent. It then relays the Registration Reply to the mobile node. The following sections describe this behavior in more detail.

If upon relaying a Registration Request to a home agent, the foreign agent receives an ICMP error message instead of a Registration Reply, then the foreign agent SHOULD send to the mobile node a Registration Reply with an appropriate "Home Agent Unreachable" failure Code (within the range 80-95, inclusive). See Section 3.7.2.3 for details on building the Registration Reply.

3.7.3.1. Validity Checks

Registration Replies with an invalid, non-zero UDP checksum MUST be silently discarded.

When a foreign agent receives a Registration Reply message, it MUST search its visitor list for a pending Registration Request with the same mobile node home address as indicated in the Reply. If no pending Request is found, the foreign agent MUST silently discard the Reply. The foreign agent MUST also silently discard the Reply if the low-order 32 bits of the Identification field in the Reply do not match those in the Request.

Also, the authentication in the Registration Reply MUST be checked. If the foreign agent and the home agent share a mobility security association, exactly one Foreign-Home Authentication Extension MUST be present in the Registration Reply, and the foreign agent MUST check the Authenticator value in the Extension. If no Foreign-Home Authentication Extension is found, or if more than one Foreign-Home Authentication Extension is found, or if the Authenticator is invalid, the foreign agent MUST silently discard the Reply and SHOULD

RED HAT 1010
379
RED HAT V. CAS

log the event as a security exception. The foreign agent also MUST reject the mobile node's registration and SHOULD send a Registration Reply to the mobile node with Code 68.

3.7.3.2. Forwarding Replies to the Mobile Node

A Registration Reply which satisfies the validity checks of Section 3.8.2.1 is relayed to the mobile node. The foreign agent MUST also update its visitor list entry for the mobile node to reflect the results of the Registration Request, as indicated by the Code field in the Reply. If the Code indicates that the mobile node has accepted the registration and the Lifetime field is nonzero, the foreign agent MUST set the Lifetime in the visitor list entry to the value specified in the Lifetime field of the Registration Reply. If, instead, the Code indicates that the Lifetime field is zero, the foreign agent MUST delete its visitor list entry for the mobile node. Finally, if the Code indicates that the registration was denied by the home agent, the foreign agent MUST delete its pending registration list entry, but not its visitor list entry, for the mobile node.

The foreign agent MUST NOT modify any of the fields beginning with the fixed portion of the Registration Reply up through and including the Mobile-Home Authentication Extension. Otherwise, an authentication failure is very likely to occur at the mobile node. In addition, the foreign agent SHOULD perform the following additional procedures:

- It MUST process and remove any Extensions following the Mobile-Home Authentication Extension,
- It MAY append its own non-authentication Extensions of relevance to the mobile node, if applicable, and
- It MUST append the Mobile-Foreign Authentication Extension, if the foreign agent shares a mobility security association with the mobile node.

Specific fields within the IP header and the UDP header of the relayed Registration Reply are set according to the same rules specified in Section 3.7.2.3.

After forwarding a valid Registration Reply to the mobile node, the foreign agent MUST update its visitor list entry for this registration as follows. If the Registration Reply indicates that the registration was accepted by the home agent, the foreign agent resets its timer of the lifetime of the registration to the Lifetime granted in the Registration Reply; unlike the mobile node's timing of the registration lifetime as described in Section 3.6.2.2, the foreign agent considers this lifetime to begin when it forwards the

RED HAT 1010
RED HAT V. CAS

Registration Reply message, ensuring that the foreign agent will not expire the registration before the mobile node does. On the other hand, if the Registration Reply indicates that the registration was rejected by the home agent, the foreign agent deletes its visitor list entry for this attempted registration.

3.8. Home Agent Considerations

Home agents play a reactive role in the registration process. The home agent receives Registration Requests from the mobile node (perhaps relayed by a foreign agent), updates its record of the mobility bindings for this mobile node, and issues a suitable Registration Reply in response to each.

A home agent MUST NOT transmit a Registration Reply except when replying to a Registration Request received from a mobile node. In particular, the home agent MUST NOT generate a Registration Reply to indicate that the Lifetime has expired.

3.8.1. Configuration and Registration Tables

Each home agent MUST be configured with an IP address and with the prefix size for the home network. The home agent MUST be configured with the home address and mobility security association of each authorized mobile node that it is serving as a home agent. When the home agent accepts a valid Registration Request from a mobile node that it serves as a home agent, the home agent MUST create or modify the entry for this mobile node in its mobility binding list containing:

- the mobile node's care-of address
- the Identification field from the Registration Reply
- the remaining Lifetime of the registration

The home agent MAY also maintain mobility security associations with various foreign agents. When receiving a Registration Request from a foreign agent, if the home agent shares a mobility security association with the foreign agent, the home agent MUST check the Authenticator in the required Foreign-Home Authentication Extension in the message, based on this mobility security association. Similarly, when sending a Registration Reply to a foreign agent, if the home agent shares a mobility security association with the foreign agent, the home agent MUST include a Foreign-Home Authentication Extension in the message, based on this mobility security association.

3.8.2. Receiving Registration Requests

RED HAT 1010
RED HAT V. CAS

If the home agent accepts an incoming Registration Request, it MUST update its record of the the mobile node's mobility binding(s) and SHOULD send a Registration Reply with a suitable Code. Otherwise (the home agent denies the Request), it SHOULD send a Registration Reply with an appropriate Code specifying the reason the Request was denied. The following sections describe this behavior in more detail.

3.8.2.1. Validity Checks

Registration Requests with an invalid, non-zero UDP checksum MUST be silently discarded by the home agent.

The authentication in the Registration Request MUST be checked. This involves the following operations:

a) The home agent MUST check for the presence of a valid Mobile-Home Authentication Extension, and perform the indicated authentication. Exactly one Mobile-Home Authentication Extension MUST be present in the Registration Request, and the home agent MUST check the Authenticator value in the Extension. If no Mobile-Home Authentication Extension is found, or if more than one Mobile-Home Authentication Extension is found, or if the Authenticator is invalid, the home agent MUST reject the mobile node's registration and SHOULD send a Registration Reply to the mobile node with Code 131. The home agent MUST then discard the Request and SHOULD log the error as a security exception.

b) The home agent MUST check that the registration Identification field is correct using the context selected by the SPI within the Mobile-Home Authentication Extension. See Section 5.6 for a description of how this is performed. If incorrect, the home agent MUST reject the Request and SHOULD send a Registration Reply to the mobile node with Code 133, including an Identification field computed in accordance with the rules specified in Section 5.6. The home agent MUST do no further processing with such a Request, though it SHOULD log the error as a security exception.

c) If the home agent shares a mobility security association with the foreign agent, the home agent MUST check for the presence of a valid Foreign-Home Authentication Extension. Exactly one Foreign-Home Authentication Extension MUST be present in the Registration Request in this case, and the home agent MUST check the Authenticator value in the Extension. If no Foreign-Home Authentication Extension is found, or if more than one Foreign-Home Authentication Extension is found, or

RED HAT 1010
RED HAT V. CAS

if the Authenticator is invalid, the home agent MUST reject the mobile node's registration and SHOULD send a Registration Reply to the mobile node with Code 132. The home agent MUST then discard the Request and SHOULD log the error as a security exception.

In addition to checking the authentication in the Registration Request, home agents MUST deny Registration Requests that are sent to the subnet-directed broadcast address of the home network (as opposed to being unicast to the home agent). The home agent MUST discard the Request and SHOULD returning a Registration Reply with a Code of 136. In this case, the Registration Reply will contain the home agent's unicast address, so that the mobile node can re-issue the Registration Request with the correct home agent address.

3.8.2.2. Accepting a Valid Request

If the Registration Request satisfies the validity checks in Section 3.8.2.1, and the home agent is able to accommodate the Request, the home agent MUST update its mobility binding list for the requesting mobile node and MUST return a Registration Reply to the mobile node. In this case, the Reply Code will be either 0 if the home agent supports simultaneous mobility bindings, or 1 if it does not. See Section 3.8.3 for details on building the Registration Reply message.

The home agent updates its record of the mobile node's mobility bindings as follows, based on the fields in the Registration Request:

- If the Lifetime is zero and the Care-of Address equals the mobile node's home address, the home agent deletes all of the entries in the mobility binding list for the requesting mobile node. This is how a mobile node requests that its home agent cease providing mobility services.

- If the Lifetime is zero and the Care-of Address does not equal the mobile node's home address, the home agent deletes only the entry containing the specified Care-of Address from the mobility binding list for the requesting mobile node. Any other active entries containing other care-of addresses will remain active.

- If the Lifetime is nonzero, the home agent adds an entry containing the requested Care-of Address to the mobility binding list for the mobile node. If the 'S' bit is set and the home agent supports simultaneous mobility bindings, the previous mobility binding entries are retained. Otherwise, the home agent removes all previous entries in the mobility binding list for the mobile node.

RED HAT 1010
RED HAT V. CAS

In all cases, the home agent MUST send a Registration Reply to the source of the Registration Request, which might indeed be a different foreign agent than that whose care-of address is being (de)registered. If the home agent shares a mobility security association with the foreign agent whose care-of address is being deregistered, and that foreign agent is different from the one which relayed the Registration Request, the home agent MAY additionally send a Registration Reply to the foreign agent whose care-of address is being deregistered. The home agent MUST NOT send such a Reply if it does not share a mobility security association with the foreign agent. If no Reply is sent, the foreign agent's visitor list will expire naturally when the original Lifetime expires.

The home agent MUST NOT increase the Lifetime above that specified by the mobile node in the Registration Request. However, it is not an error for the mobile node to request a Lifetime longer than the home agent is willing to accept. In this case, the home agent simply reduces the Lifetime to a permissible value and returns this value in the Registration Reply. The Lifetime value in the Registration Reply informs the mobile node of the granted lifetime of the registration, indicating when it SHOULD re-register in order to maintain continued service. After the expiration of this registration lifetime, the home agent MUST delete its entry for this registration in its mobility binding list.

If the Registration Request duplicates an accepted current Registration Request, the new Lifetime MUST NOT extend beyond the Lifetime originally granted. A Registration Request is a duplicate if the home address, care-of address, and Identification fields all equal those of an accepted current registration.

In addition, if the home network implements ARP [16], and the Registration Request asks the home agent to create a mobility binding for a mobile node which previously had no binding (the mobile node was previously assumed to be at home), then the home agent MUST follow the procedures described in Section 4.6 with regard to ARP, proxy ARP, and gratuitous ARP. If the mobile node already had a previous mobility binding, the home agent MUST continue to follow the rules for proxy ARP described in Section 4.6.

3.8.2.3. Denying an Invalid Request

If the Registration Reply does not satisfy all of the validity checks in Section 3.8.2.1, or the home agent is unable to accommodate the Request, the home agent SHOULD return a Registration Reply to the mobile node with a Code that indicates the reason for the error. If a foreign agent was involved in relaying the Request, this allows the foreign agent to delete its pending visitor list entry. Also, this

RED HAT 1010
RED HAT V. CAS

informs the mobile node of the reason for the error such that it may attempt to fix the error and issue another Request.

This section lists a number of reasons the home agent might reject a Request, and provides the Code value it should use in each instance. See Section 3.8.3 for additional details on building the Registration Reply message.

Many reasons for rejecting a registration are administrative in nature. For example, a home agent can limit the number of simultaneous registrations for a mobile node, by rejecting any registrations that would cause its limit to be exceeded, and returning a Registration Reply with error code 135. Similarly, a home agent may refuse to grant service to mobile nodes which have entered unauthorized service areas by returning a Registration Reply with a Code of 129.

If the Reserved field is nonzero, it MUST deny the Request with a Code of 134.

3.8.3. Sending Registration Replies

If the home agent accepts a Registration Request, it then MUST update its record of the mobile node's mobility binding(s) and SHOULD send a Registration Reply with a suitable Code. Otherwise (the home agent has denied the Request), it SHOULD send a Registration Reply with an appropriate Code specifying the reason the Request was denied. The following sections provide additional detail for the values the home agent MUST supply in the fields of Registration Reply messages.

3.8.3.1. IP/UDP Fields

This section provides the specific rules by which mobile nodes pick values for the IP and UDP header fields of a Registration Reply.

IP Source Address
: Copied from the IP Destination Address of Registration Request, unless a multicast or broadcast address was used. If the IP Destination Address of the Registration Request was a broadcast or multicast address, the IP Source Address of the Registration Reply MUST be set to the home agent's (unicast) IP address.

IP Destination Address
: Copied from the IP Source Address of the Registration Request.

RED HAT 1010
RED HAT V. CAS

UDP Source Port
Copied from the UDP Destination Port of the Registration Request.

UDP Destination Port
Copied from the UDP Source Port of the Registration Request.

When sending a Registration Reply in response to a Registration Request that requested deregistration of the mobile node (the Lifetime is zero and the Care-of Address equals the mobile node's home address) and in which the IP Source Address was also set to the mobile node's home address (this is the normal method used by a mobile node to deregister when it returns to its home network), the IP Destination Address in the Registration Reply will be set to the mobile node's home address, as copied from the IP Source Address of the Request.

In this case, when transmitting the Registration Reply, the home agent MUST transmit the Reply directly onto the home network as if the mobile node were at home, bypassing any mobility binding list entry that may still exist at the home agent for the destination mobile node. In particular, for a mobile node returning home after being registered with a care-of address, if the mobile node's new Registration Request is not accepted by the home agent, the mobility binding list entry for the mobile node will still indicate that datagrams addressed to the mobile node should be tunneled to the mobile node's registered care-of address; when sending the Registration Reply indicating the rejection of this Request, this existing binding list entry MUST be ignored, and the home agent MUST transmit this Reply as if the mobile node were at home.

3.8.3.2. Registration Reply Fields

This section provides specific rules by which home agents pick values for the fields within the fixed portion of a Registration Reply. The Code field of the Registration Reply is chosen in accordance with the rules specified in the previous sections. When replying to an accepted registration, a home agent SHOULD respond with Code 1 if it does not support simultaneous registrations.

The Lifetime field MUST be copied from the corresponding field in the Registration Request, unless the requested value is greater than the maximum length of time the home agent is willing to provide the requested service. In such a case, the Lifetime MUST be set to the length of time that service will actually be provided by the home agent. This reduced Lifetime SHOULD be the maximum Lifetime allowed by the home agent (for this mobile node and care-of address).

RED HAT 1010
RED HAT V. CAS

The Home Address field MUST be copied from the corresponding field in the Registration Request.

If the Home Agent field in the Registration Request contains a unicast address of this home agent, then that field MUST be copied into the Home Agent field of the Registration Reply. Otherwise, the home agent MUST set the Home Agent field in the Registration Reply to its unicast address. In this latter case, the home agent MUST reject the registration with a suitable code (e.g., Code 136) to prevent the mobile node from possibly being simultaneously registered with two or more home agents.

3.8.3.3. Extensions

This section describes the ordering of any required and any optional Mobile IP Extensions that a home agent appends to a Registration Reply. The following ordering MUST be followed:

a) The IP header, followed by the UDP header, followed by the fixed-length portion of the Registration Reply,

b) If present, any non-authentication Extensions used by the mobile node (which may or may not also be used by the foreign agent),

c) The Mobile-Home Authentication Extension,

d) If present, any non-authentication Extensions used only by the foreign agent, and

e) The Foreign-Home Authentication Extension, if present.

Note that items (a) and (c) MUST appear in every Registration Reply sent by the home agent. Items (b), (d), and (e) are optional. However, item (e) MUST be included when the home agent and the foreign agent share a mobility security association.

4. Routing Considerations

This section describes how mobile nodes, home agents, and (possibly) foreign agents cooperate to route datagrams to/from mobile nodes that are connected to a foreign network. The mobile node informs its home agent of its current location using the registration procedure described in Section 3. See the protocol overview in Section 1.7 for the relative locations of the mobile node's home address with respect to its home agent, and the mobile node itself with respect to any foreign agent with which it might attempt to register.

RED HAT 1010
RED HAT V. CAS

4.1. Encapsulation Types

Home agents and foreign agents MUST support tunneling datagrams using IP in IP encapsulation [14]. Any mobile node that uses a co-located care-of address MUST support receiving datagrams tunneled using IP in IP encapsulation. Minimal encapsulation [15] and GRE encapsulation [8] are alternate encapsulation methods which MAY optionally be supported by mobility agents and mobile nodes. The use of these alternative forms of encapsulation, when requested by the mobile node, is otherwise at the discretion of the home agent.

4.2. Unicast Datagram Routing

4.2.1. Mobile Node Considerations

When connected to its home network, a mobile node operates without the support of mobility services. That is, it operates in the same way as any other (fixed) host or router. The method by which a mobile node selects a default router when connected to its home network, or when away from home and using a co-located care-of address, is outside the scope of this document. ICMP Router Advertisement [4] is one such method.

When registered on a foreign network, the mobile node chooses a default router by the following rules:

- If the mobile node is registered using a foreign agent care-of address, then the mobile node MUST choose its default router from among the Router Addresses advertised in the ICMP Router Advertisement portion of that Agent Advertisement message. The mobile node MAY also consider the IP source address of the Agent Advertisement as another possible choice for the IP address of a default router, along with the (possibly empty) list of Router Addresses from the ICMP Router Advertisement portion of the message. In such cases, the IP source address MUST be considered to be the worst choice (lowest preference) for a default router.

- If the mobile node is registered directly with its home agent using a co-located care-of address, then the mobile node SHOULD choose its default router from among those advertised in any ICMP Router Advertisement message that it receives for which its externally obtained care-of address and the Router Address match under the network prefix. If the mobile node's externally obtained care-of address matches the IP source address of the Agent Advertisement under the network prefix, the mobile node MAY also consider that IP source address as another possible choice for the IP address of a default router, along with the (possibly empty) list of Router Addresses from the ICMP Router

RED HAT 1010
RED HAT V. CAS

Advertisement portion of the message. If so, the IP source address MUST be considered to be the worst choice (lowest preference) for a default router. The network prefix MAY be obtained from the Prefix-Lengths Extension in the Router Advertisement, if present. The prefix MAY also be obtained through other mechanisms beyond the scope of this document.

Beyond these rules, the actual selection of the default router is made by the selection method specified for ICMP Router Discovery [4], among the Router Addresses specified above. In any case, a mobile node registered via a foreign agent MAY choose its foreign agent as a default router.

Note that Van Jacobson header compression [10] will not function properly unless all TCP IP datagrams to and from the mobile node pass, respectively, through the same first and last-hop router. The mobile node, therefore, MUST select its foreign agent as its default router if it performs Van Jacobson header compression with its foreign agent.

4.2.2. Foreign Agent Considerations

Upon receipt of an encapsulated datagram sent to its advertised care-of address, a foreign agent MUST compare the inner destination address to those entries in its visitor list. When the destination does not match the address of any mobile node currently in the visitor list, the foreign agent MUST NOT forward the datagram without modifications to the original IP header, because otherwise a routing loop is likely to result. The datagram SHOULD be silently discarded. ICMP Destination Unreachable MUST NOT be sent when a foreign agent is unable to forward an incoming tunneled datagram. Otherwise, the foreign agent forwards the decapsulated datagram to the mobile node.

The foreign agent MUST NOT advertise to other routers in its routing domain, nor to any other mobile node, the presence of a mobile router (Section 4.5).

The foreign agent MUST route datagrams it receives from registered mobile nodes. At a minimum, this means that the foreign agent must verify the IP Header Checksum, decrement the IP Time To Live, recompute the IP Header Checksum, and forward such datagrams to a default router. In addition, the foreign agent SHOULD send an appropriate ICMP Redirect message to the mobile node.

RED HAT 1010
RED HAT V. CAS

4.2.3. Home Agent Considerations

The home agent MUST be able to intercept any datagrams on the home network addressed to the mobile node while the mobile node is registered away from home. Proxy and gratuitous ARP MAY be used in enabling this interception, as specified in Section 4.6.

The home agent must examine the IP Destination Address of all arriving datagrams to see if it is equal to the home address of any of its mobile nodes registered away from home. If so, the home agent tunnels the datagram to the mobile node's currently registered care-of address or addresses. If the home agent supports the optional capability of multiple simultaneous mobility bindings, it tunnels a copy to each care-of address in the mobile node's mobility binding list. If the mobile node has no current mobility bindings, the home agent MUST NOT attempt to intercept datagrams destined for the mobile node, and thus will not in general receive such datagrams. However, if the home agent is also a router handling common IP traffic, it is possible that it will receive such datagrams for forwarding onto the home network. In this case, the home agent MUST assume the mobile node is at home and simply forward the datagram directly onto the home network.

See Section 4.1 regarding methods of encapsulation that may be used for tunneling. Nodes implementing tunneling SHOULD also implement the "tunnel soft state" mechanism [14], which allows ICMP error messages returned from the tunnel to correctly be reflected back to the original senders of the tunneled datagrams.

Home agents SHOULD be able to decapsulate and further deliver packets addressed to themselves, sent by a mobile node for the purpose of maintaining location privacy, as described in Section 5.5.

If the Lifetime for a given mobility binding expires before the home agent has received another valid Registration Request for that mobile node, then that binding is deleted from the mobility binding list. The home agent MUST NOT send any Registration Reply message simply because the mobile node's binding has expired. The entry in the visitor list of the mobile node's current foreign agent will expire naturally, probably at the same time as the binding expired at the home agent. When a mobility binding's lifetime expires, the home agent MUST delete the binding, but it MUST retain any other (non-expired) simultaneous mobility bindings that it holds for the mobile node.

When a home agent receives a datagram, intercepted for one of its mobile nodes registered away from home, the home agent MUST examine the datagram to check if it is already encapsulated. If so, special

RED HAT 1010
RED HAT V. CAS

rules apply in the forwarding of that datagram to the mobile node:

- If the inner (encapsulated) Destination Address is the same as the outer Destination Address (the mobile node), then the home agent MUST also examine the outer Source Address of the encapsulated datagram (the source address of the tunnel). If this outer Source Address is the same as the mobile node's current care-of address, the home agent MUST silently discard that datagram in order to prevent a likely routing loop. If, instead, the outer Source Address is NOT the same as the mobile node's current care-of address, then the home agent SHOULD forward the datagram to the mobile node. In order to forward the datagram in this case, the home agent MAY simply alter the outer Destination Address to the care-of address, rather than re-encapsulating the datagram.

- Otherwise (the inner Destination Address is NOT the same as the outer Destination Address), the home agent SHOULD encapsulate the datagram again (recursive encapsulation), with the new outer Destination Address set equal to the mobile node's care-of address. That is, the home agent forwards the entire datagram to the mobile node in the same way as any other datagram (encapsulated already or not).

4.3. Broadcast Datagrams

When a home agent receives a broadcast datagram, it MUST NOT forward the datagram to any mobile nodes in its mobility binding list other than those that have requested forwarding of broadcast datagrams. A mobile node MAY request forwarding of broadcast datagrams by setting the 'B' bit in its Registration Request message (Section 3.3). For each such registered mobile node, the home agent SHOULD forward received broadcast datagrams to the mobile node, although it is a matter of configuration at the home agent as to which specific categories of broadcast datagrams will be forwarded to such mobile nodes.

If the 'D' bit was set in the mobile node's Registration Request message, indicating that the mobile node is using a co-located care-of address, the home agent simply tunnels appropriate broadcast IP datagrams to the mobile node's care-of address. Otherwise (the 'D' bit was NOT set), the home agent first encapsulates the broadcast datagram in a unicast datagram addressed to the mobile node's home address, and then tunnels this encapsulated datagram to the foreign agent. This extra level of encapsulation is required so that the foreign agent can determine which mobile node should receive the datagram after it is decapsulated. When received by the foreign agent, the unicast encapsulated datagram is detunneled and delivered

RED HAT 1010
RED HAT V. CAS

to the mobile node in the same way as any other datagram. In either case, the mobile node must decapsulate the datagram it receives in order to recover the original broadcast datagram.

4.4. Multicast Datagram Routing

As mentioned previously, a mobile node that is connected to its home network functions in the same way as any other (fixed) host or router. Thus, when it is at home, a mobile node functions identically to other multicast senders and receivers. This section therefore describes the behavior of a mobile node that is visiting a foreign network.

In order receive multicasts, a mobile node MUST join the multicast group in one of two ways. First, a mobile node MAY join the group via a (local) multicast router on the visited subnet. This option assumes that there is a multicast router present on the visited subnet. If the mobile node is using a co-located care-of address, it SHOULD use this address as the source IP address of its IGMP [5] messages. Otherwise, it MUST use its home address.

Alternatively, a mobile node which wishes to receive multicasts MAY join groups via a bi-directional tunnel to its home agent, assuming that its home agent is a multicast router. The mobile node tunnels IGMP messages to its home agent and the home agent forwards multicast datagrams down the tunnel to the mobile node. The rules for multicast datagram delivery to mobile nodes in this case are identical to those for broadcast datagrams (Section 4.3). Namely, if the mobile node is using a co-located care-of address (the 'D' bit was set in the mobile node's Registration Request), then the home agent SHOULD tunnel the datagram to this care-of address; otherwise, the home agent MUST first encapsulate the datagram in a unicast datagram addressed to the mobile node's home address and then MUST tunnel the resulting datagram (recursive tunneling) to the mobile node's care-of address.

A mobile node that wishes to send datagrams to a multicast group also has two options: (1) send directly on the visited network; or (2) send via a tunnel to its home agent. Because multicast routing in general depends upon the IP source address, a mobile node which sends multicast datagrams directly on the visited network MUST use a co-located care-of address as the IP source address. Similarly, a mobile node which tunnels a multicast datagram to its home agent MUST use its home address as the IP source address of both the (inner) multicast datagram and the (outer) encapsulating datagram. This second option assumes that the home agent is a multicast router.

RED HAT 1010
RED HAT V. CAS

4.5. Mobile Routers

A mobile node can be a router, which is responsible for the mobility of one or more entire networks moving together, perhaps on an airplane, a ship, a train, an automobile, a bicycle, or a kayak. The nodes connected to a network served by the mobile router may themselves be fixed nodes or mobile nodes or routers. In this document, such networks are called "mobile networks".

A mobile router MAY act as a foreign agent and provide a foreign agent care-of address to mobile nodes connected to the mobile network. Typical routing to a mobile node via a mobile router in this case is illustrated by the following example:

a) A laptop computer is disconnected from its home network and later attached to a network port in the seat back of an aircraft. The laptop computer uses Mobile IP to register on this foreign network, using a foreign agent care-of address discovered through an Agent Advertisement from the aircraft's foreign agent.

b) The aircraft network is itself mobile. Suppose the node serving as the foreign agent on the aircraft also serves as the default router that connects the aircraft network to the rest of the Internet. When the aircraft is at home, this router is attached to some fixed network at the airline's headquarters, which is the router's home network. While the aircraft is in flight, this router registers from time to time over its radio link with a series of foreign agents below it on the ground. This router's home agent is a node on the fixed network at the airline's headquarters.

c) Some correspondent node sends a datagram to the laptop computer, addressing the datagram to the laptop's home address. This datagram is initially routed to the laptop's home network.

d) The laptop's home agent intercepts the datagram on the home network and tunnels it to the laptop's care-of address, which in this example is an address of the node serving as router and foreign agent on the aircraft. Normal IP routing will route the datagram to the fixed network at the airline's headquarters.

RED HAT 1010
RED HAT V. CAS

e) The aircraft router and foreign agent's home agent there intercepts the datagram and tunnels it to its current care-of address, which in this example is some foreign agent on the ground below the aircraft. The original datagram from the correspondent node has now been encapsulated twice: once by the laptop's home agent and again by the aircraft's home agent.

f) The foreign agent on the ground decapsulates the datagram, yielding a datagram still encapsulated by the laptop's home agent, with a destination address of the laptop's care-of address. The ground foreign agent sends the resulting datagram over its radio link to the aircraft.

g) The foreign agent on the aircraft decapsulates the datagram, yielding the original datagram from the correspondent node, with a destination address of the laptop's home address. The aircraft foreign agent delivers the datagram over the aircraft network to the laptop's link-layer address.

This example illustrated the case in which a mobile node is attached to a mobile network. That is, the mobile node is mobile with respect to the network, which itself is also mobile (here with respect to the ground). If, instead, the node is fixed with respect to the mobile network (the mobile network is the fixed node's home network), then either of two methods may be used to cause datagrams from correspondent nodes to be routed to the fixed node.

A home agent MAY be configured to have a permanent registration for the fixed node, that indicates the mobile router's address as the fixed host's care-of address. The mobile router's home agent will usually be used for this purpose. The home agent is then responsible for advertising connectivity using normal routing protocols to the fixed node. Any datagrams sent to the fixed node will thus use recursive tunneling as described above.

Alternatively, the mobile router MAY advertise connectivity to the entire mobile network using normal IP routing protocols through a bi-directional tunnel to its own home agent. This method avoids the need for recursive tunneling of datagrams.

4.6. ARP, Proxy ARP, and Gratuitous ARP

The use of ARP [16] requires special rules for correct operation when wireless or mobile nodes are involved. The requirements specified in this section apply to all home networks in which ARP is used for address resolution.

RED HAT 1010
RED HAT V. CAS

In addition to the normal use of ARP for resolving a target node's link-layer address from its IP address, this document distinguishes two special uses of ARP:

- A Proxy ARP [18] is an ARP Reply sent by one node on behalf of another node which is either unable or unwilling to answer its own ARP Requests. The sender of a Proxy ARP reverses the Sender and Target Protocol Address fields as described in [16], but supplies some configured link-layer address (generally, its own) in the Sender Hardware Address field. The node receiving the Reply will then associate this link-layer address with the IP address of the original target node, causing it to transmit future datagrams for this target node to the node with that link-layer address.

- A Gratuitous ARP [23] is an ARP packet sent by a node in order to spontaneously cause other nodes to update an entry in their ARP cache. A gratuitous ARP MAY use either an ARP Request or an ARP Reply packet. In either case, the ARP Sender Protocol Address and ARP Target Protocol Address are both set to the IP address of the cache entry to be updated, and the ARP Sender Hardware Address is set to the link-layer address to which this cache entry should be updated. When using an ARP Reply packet, the Target Hardware Address is also set to the link-layer address to which this cache entry should be updated (this field is not used in an ARP Request packet).

  In either case, for a gratuitous ARP, the ARP packet MUST be transmitted as a local broadcast packet on the local link. As specified in [16], any node receiving any ARP packet (Request or Reply) MUST update its local ARP cache with the Sender Protocol and Hardware Addresses in the ARP packet, if the receiving node has an entry for that IP address already in its ARP cache. This requirement in the ARP protocol applies even for ARP Request packets, and for ARP Reply packets that do not match any ARP Request transmitted by the receiving node [16].

While a mobile node is registered on a foreign network, its home agent uses proxy ARP [18] to reply to ARP Requests it receives that seek the mobile node's link-layer address. When receiving an ARP Request, the home agent MUST examine the target IP address of the Request, and if this IP address matches the home address of any mobile node for which it has a registered mobility binding, the home agent MUST transmit an ARP Reply on behalf of the mobile node. After exchanging the sender and target addresses in the packet [18], the home agent MUST set the sender link-layer address in the packet to the link-layer address of its own interface over which the Reply will be sent.

RED HAT 1010
RED HAT V. CAS

When a mobile node leaves its home network and registers a binding on a foreign network, its home agent uses gratuitous ARP to update the ARP caches of nodes on the home network. This causes such nodes to associate the link-layer address of the home agent with the mobile node's home (IP) address. When registering a binding for a mobile node for which the home agent previously had no binding (the mobile node was assumed to be at home), the home agent MUST transmit a gratuitous ARP on behalf of the mobile node. This gratuitous ARP packet MUST be transmitted as a broadcast packet on the link on which the mobile node's home address is located. Since broadcasts on the local link (such as Ethernet) are typically not guaranteed to be reliable, the gratuitous ARP packet SHOULD be retransmitted a small number of times to increase its reliability.

When a mobile node returns to its home network, the mobile node and its home agent use gratuitous ARP to cause all nodes on the mobile node's home network to update their ARP caches to once again associate the mobile node's own link-layer address with the mobile node's home (IP) address. Before transmitting the (de)Registration Request message to its home agent, the mobile node MUST transmit this gratuitous ARP on its home network as a local broadcast on this link. The gratuitous ARP packet SHOULD be retransmitted a small number of times to increase its reliability, but these retransmissions SHOULD proceed in parallel with the transmission and processing of its (de)Registration Request.

When the mobile node's home agent receives and accepts this (de)Registration Request, the home agent MUST also transmit a gratuitous ARP on the mobile node's home network. This gratuitous ARP also is used to associate the mobile node's home address with the mobile node's own link-layer address. A gratuitous ARP is transmitted by both the mobile node and its home agent, since in the case of wireless network interfaces, the area within transmission range of the mobile node will likely differ from that within range of its its home agent. Th ARP packet from the home agent MUST be transmitted as a local broadcast on the mobile node's home link, and SHOULD be retransmitted a small number of times to increase its reliability; these retransmissions, however, SHOULD proceed in parallel with the transmission and processing of its (de)Registration Reply.

While the mobile node is away from home, it MUST NOT transmit any broadcast ARP Request or ARP Reply messages. Finally, while the mobile node is away from home, it MUST NOT reply to ARP Requests in which the target IP address is its own home address, unless the ARP Request is sent by a foreign agent with which the mobile node is attempting to register or a foreign agent with which the mobile node has an unexpired registration. In the latter case, the mobile

RED HAT 1010
RED HAT V. CAS

node MUST use a unicast ARP Reply to respond to the foreign agent. Note that if the mobile node is using a co-located care-of address and receives an ARP Request in which the target IP address is this care-of address, then the mobile node SHOULD reply to this ARP Request. Note also that, when transmitting a Registration Request on a foreign network, a mobile node may discover the link-layer address of a foreign agent by storing the address as it is received from the Agent Advertisement from that foreign agent, but not by transmitting a broadcast ARP Request message.

The specific order in which each of the above requirements for the use of ARP, proxy ARP, and gratuitous ARP are applied, relative to the transmission and processing of the mobile node's Registration Request and Registration Reply messages when leaving home or returning home, are important to the correct operation of the protocol.

To summarize the above requirements, when a mobile node leaves its home network, the following steps, in this order, MUST be performed:

- The mobile node decides to register away from home, perhaps because it has received an Agent Advertisement from a foreign agent and has not recently received one from its home agent.

- Before transmitting the Registration Request, the mobile node disables its own future processing of any ARP Requests it may subsequently receive requesting the link-layer address corresponding to its home address, except insofar as necessary to communicate with foreign agents on visited networks.

- The mobile node transmits its Registration Request.

- When the mobile node's home agent receives and accepts the Registration Request, it performs a gratuitous ARP on behalf of the mobile node, and begins using proxy ARP to reply to ARP Requests that it receives requesting the mobile node's link-layer address. If, instead, the home agent rejects the Registration Request, no ARP processing (gratuitous nor proxy) is performed by the home agent.

When a mobile node later returns to its home network, the following steps, in this order, MUST be performed:

- The mobile node decides to register at home, perhaps because it has received an Agent Advertisement from its home agent.

RED HAT 1010
RED HAT V. CAS

- Before transmitting the Registration Request, the mobile node re-enables its own future processing of any ARP Requests it may subsequently receive requesting its link-layer address.

- The mobile node performs a gratuitous ARP for itself.

- The mobile node transmits its Registration Request.

- When the mobile node's home agent receives and accepts the Registration Request, it stops using proxy ARP to reply to ARP Requests that it receives requesting the mobile node's link-layer address, and then performs a gratuitous ARP on behalf of the mobile node. If, instead, the home agent rejects the Registration Request, no ARP processing (gratuitous nor proxy) is performed by the home agent.

5. Security Considerations

The mobile computing environment is potentially very different from the ordinary computing environment. In many cases, mobile computers will be connected to the network via wireless links. Such links are particularly vulnerable to passive eavesdropping, active replay attacks, and other active attacks.

5.1. Message Authentication Codes

Home agents and mobile nodes MUST be able to perform authentication. The default algorithm is keyed MD5 [21], with a key size of 128 bits. The default mode of operation is to both precede and follow the data to be hashed, by the 128-bit key; that is, MD5 is to be used in "prefix+suffix" mode. The foreign agent MUST also support authentication using keyed MD5 and key sizes of 128 bits or greater, with manual key distribution. More authentication algorithms, algorithm modes, key distribution methods, and key sizes MAY also be supported.

5.2. Areas of Security Concern in this Protocol

The registration protocol described in this document will result in a mobile node's traffic being tunneled to its care-of address. This tunneling feature could be a significant vulnerability if the registration were not authenticated. Such remote redirection, for instance as performed by the mobile registration protocol, is widely understood to be a security problem in the current Internet if not authenticated [2]. Moreover, the Address Resolution Protocol (ARP) is not authenticated, and can potentially be used to steal another host's traffic. The use of "Gratuitous ARP" (Section 4.6) brings with it all of the risks associated with the use of ARP.

RED HAT 1010
RED HAT V. CAS

5.3. Key Management

This specification requires a strong authentication mechanism (keyed MD5) which precludes many potential attacks based on the Mobile IP registration protocol. However, because key distribution is difficult in the absence of a network key management protocol, messages with the foreign agent are not all required to be authenticated. In a commercial environment it might be important to authenticate all messages between the foreign agent and the home agent, so that billing is possible, and service providers do not provide service to users that are not legitimate customers of that service provider.

5.4. Picking Good Random Numbers

The strength of any authentication mechanism depends on several factors, including the innate strength of the authentication algorithm, the secrecy of the key used, the strength of the key used, and the quality of the particular implementation. This specification requires implementation of keyed MD5 for authentication, but does not preclude the use of other authentication algorithms and modes. For keyed MD5 authentication to be useful, the 128-bit key must be both secret (that is, known only to authorized parties) and pseudo-random. If nonces are used in connection with replay protection, they must also be selected carefully. Eastlake, et al. [7] provides more information on generating pseudo-random numbers.

5.5. Privacy

Users who have sensitive data that they do not wish others to see should use mechanisms outside the scope of this document (such as encryption) to provide appropriate protection. Users concerned about traffic analysis should consider appropriate use of link encryption. If absolute location privacy is desired, the mobile node can create a tunnel to its home agent. Then, datagrams destined for correspondent nodes will appear to emanate from the home network, and it may be more difficult to pinpoint the location of the mobile node. Such mechanisms are all beyond the scope of this document.

RED HAT 1010
RED HAT V. CAS

5.6. Replay Protection for Registration Requests

The Identification field is used to let the home agent verify that a registration message has been freshly generated by the mobile node, not replayed by an attacker from some previous registration. Two methods are described in this section: timestamps (mandatory) and "nonces" (optional). All mobile nodes and home agents MUST implement timestamp-based replay protection. These nodes MAY also implement nonce-based replay protection (but see Appendix A.2 for a patent that may apply to nonce-based replay protection).

The style of replay protection in effect between a mobile node and its home agent is part of the mobile security association. A mobile node and its home agent MUST agree on which method of replay protection will be used. The interpretation of the Identification field depends on the method of replay protection as described in the subsequent subsections.

Whatever method is used, the low-order 32 bits of the Identification MUST be copied unchanged from the Registration Request to the Reply. The foreign agent uses those bits (and the mobile node's home address) to match Registration Requests with corresponding replies. The mobile node MUST verify that the low-order 32 bits of any Registration Reply are identical to the bits it sent in the Registration Request.

The Identification in a new Registration Request MUST NOT be the same as in an immediately preceding Request, and SHOULD NOT repeat while the same security context is being used between the mobile node and the home agent. Retransmission as in Section 3.6.3 is allowed.

5.6.1. Replay Protection using Timestamps

The basic principle of timestamp replay protection is that the node generating a message inserts the current time of day, and the node receiving the message checks that this timestamp is sufficiently close to its own time of day. Obviously the two nodes must have adequately synchronized time-of-day clocks. As with any messages, time synchronization messages may be protected against tampering by an authentication mechanism determined by the security context between the two nodes.

If timestamps are used, the mobile node MUST set the Identification field to a 64-bit value formatted as specified by the Network Time Protocol [13]. The low-order 32 bits of the NTP format represent fractional seconds, and those bits which are not available from a time source SHOULD be generated from a good source of randomness. Note, however, that when using timestamps, the 64-bit Identification

RED HAT 1010
RED HAT V. CAS

used in a Registration Request from the mobile node MUST be greater than that used in any previous Registration Request, as the home agent uses this field also as a sequence number. Without such a sequence number, it would be possible for a delayed duplicate of an earlier Registration Request to arrive at the home agent (within the clock synchronization required by the home agent), and thus be applied out of order, mistakenly altering the mobile node's current registered care-of address.

Upon receipt of a Registration Request with a valid Mobile-Home Authentication Extension, the home agent MUST check the Identification field for validity. In order to be valid, the timestamp contained in the Identification field MUST be close enough to the home agent's time of day clock and the timestamp MUST be greater than all previously accepted timestamps for the requesting mobile node. Time tolerances and resynchronization details are specific to a particular mobility security association.

If the timestamp is valid, the home agent copies the entire Identification field into the Registration Reply it returns the Reply to the mobile node. If the timestamp is not valid, the home agent copies only the low-order 32 bits into the Registration Reply, and supplies the high-order 32 bits from its own time of day. In this latter case, the home agent MUST reject the registration by returning Code 133 (identification mismatch) in the Registration Reply.

As described in Section 3.6.2.1, the mobile node MUST verify that the low-order 32 bits of the Identification in the Registration Reply are identical to those in the rejected registration attempt, before using the high-order bits for clock resynchronization.

5.6.2. Replay Protection using Nonces

Implementors of this optional mechanism should examine Appendix A.2 for a patent that may be applicable to nonce-based replay protection.

The basic principle of nonce replay protection is that node A includes a new random number in every message to node B, and checks that node B returns that same number in its next message to node A. Both messages use an authentication code to protect against alteration by an attacker. At the same time node B can send its own nonces in all messages to node A (to be echoed by node A), so that it too can verify that it is receiving fresh messages.

The home agent may be expected to have resources for computing pseudo-random numbers useful as nonces [7]. It inserts a new nonce as the high-order 32 bits of the identification field of every Registration Reply. The home agent copies the low-order 32 bits of

RED HAT 1010
RED HAT V. CAS

the Identification from the Registration Request message into the low-order 32 bits of the Identification in the Registration Reply. When the mobile node receives an authenticated Registration Reply from the home agent, it saves the high-order 32 bits of the identification for use as the high-order 32 bits of its next Registration Request.

The mobile node is responsible for generating the low-order 32 bits of the Identification in each Registration Request. Ideally it should generate its own random nonces. However it may use any expedient method, including duplication of the random value sent by the home agent. The method chosen is of concern only to the mobile node, because it is the node that checks for valid values in the Registration Reply. The high-order and low-order 32 bits of the identification chosen SHOULD both differ from their previous values. The home agent uses a new high-order value and the mobile node uses a new low-order value for each registration message. The foreign agent uses the low-order value (and the mobile host's home address) to correctly match registration replies with pending Requests (Section 3.7.1).

If a registration message is rejected because of an invalid nonce, the Reply always provides the mobile node with a new nonce to be used in the next registration. Thus the nonce protocol is self-synchronizing.

6. Acknowledgments

Special thanks to Steve Deering (Xerox PARC), along with Dan Duchamp and John Ioannidis (JI) (Columbia), for forming the working group, chairing it, and putting so much effort into its early development.

Thanks also to Kannan Alaggapan, Greg Minshall, and Tony Li for their contributions to the group while performing the duties of chairperson, as well as for their many useful comments.

Thanks to the active members of the Mobile IP Working Group, particularly those who contributed text, including (in alphabetical order)

- Ran Atkinson (Naval Research Lab),
- Dave Johnson (Carnegie Mellon University),
- Frank Kastenholz (FTP Software),
- Anders Klemets (KTH),
- Chip Maguire (KTH),
- Andrew Myles (Macquarie University),
- Al Quirt (Bell Northern Research),
- Yakov Rekhter (IBM), and
- Fumio Teraoka (Sony).

Thanks to Charlie Kunzinger and to Bill Simpson, the editors who produced the first drafts for of this document, reflecting the discussions of the Working Group. Much of the new text of this memo is due to Jim Solomon and Dave Johnson.

Thanks to Greg Minshall (Novell), Phil Karn (Qualcomm), and Frank Kastenholz (FTP Software) for their generous support in hosting interim Working Group meetings.

RED HAT 1010
RED HAT V. CAS

A. Patent Issues

As of the time of publication, the IETF had been made aware of two patents that may be relevant to implementors of the protocol described in this technical specification.

A.1. IBM Patent #5,159,592

Charles Perkins, editor of this memo, is sole inventor of U.S. Patent No. 5,159,592, assigned to IBM. In a letter dated May 30, 1995, IBM brought this patent to the attention of the IETF, stating that this patent "relates to the Mobile IP." We understand that IBM did not intend to assert that any particular implementation of Mobile IP would or would not infringe the patent, but rather that IBM was meeting what it viewed as a duty to disclose information that could be relevant to the process of adopting a standard.

Based on a review of the claims of the patent, IETF believes that a system of registering an address obtained from a foreign agent, as described in the document, would not necessarily infringe any of the claims of the patent; and that a system in which an address is obtained elsewhere and then registered can be implemented without necessarily infringing any claims of the patent. Accordingly, our view is that the proposed protocol can be implemented without necessarily infringing the Perkins Patent.

Parties considering adopting this protocol must be aware that some specific implementations, or features added to otherwise non-infringing implementations, may raise an issue of infringement with respect to this patent or to some other patent.

This statement is for the IETF's assistance in its standard-setting procedure, and should not be relied upon by any party as an opinion or guarantee that any implementation it might make or use would not be covered by the Perkins Patent and any other patents. In particular, IBM might disagree with the interpretation of this patent described herein.

A.2. IBM Patent #5,148,479

This patent, also assigned to IBM, may be relevant to those who implement nonce-based replay protection as described in Section 5.6.2. Note that nonce-based replay protection is an optional feature of this specification. Timestamp-based replay protection, on the other hand, (Section 5.6.1) is a requirement of this specification.

RED HAT 1010
RED HAT V. CAS

B. Link-Layer Considerations

The mobile node MAY use link-layer mechanisms to decide that its point of attachment has changed. Such indications include the Down/Testing/Up interface status [11], and changes in cell or administration. The mechanisms will be specific to the particular link-layer technology, and are outside the scope of this document.

The Point-to-Point-Protocol (PPP) [22] and its Internet Protocol Control Protocol (IPCP) [12], negotiates the use of IP addresses.

The mobile node SHOULD first attempt to specify its home address, so that if the mobile node is attaching to its home network, the unrouted link will function correctly. When the home address is not accepted by the peer, but a transient IP address is dynamically assigned to the mobile node, and the mobile node is capable of supporting a co-located care-of address, the mobile node MAY register that address as a co-located care-of address. When the peer specifies its own IP address, that address MUST NOT be assumed to be a foreign agent care-of address or the IP address of a home agent.

C. TCP Considerations

C.1. TCP Timers

Most hosts and routers which implement TCP/IP do not permit easy configuration of the TCP timer values. When high-delay (e.g., SATCOM) or low-bandwidth (e.g., High-Frequency Radio) links are in use, the default TCP timer values in many systems may cause retransmissions or timeouts, even when the link and network are actually operating properly with greater than usual delays because of the medium in use. This can cause an inability to create or maintain TCP connections over such links, and can also cause unneeded retransmissions which consume already scarce bandwidth. Vendors are encouraged to make TCP timers more configurable. Vendors of systems designed for the mobile computing markets should pick default timer values more suited to low-bandwidth, high-delay links. Users of mobile nodes should be sensitive to the possibility of timer-related difficulties.

C.2. TCP Congestion Management

Mobile nodes often use media which are more likely to introduce errors, effectively causing more packets to be dropped. This introduces a conflict with the mechanisms for congestion management found in modern versions of TCP [9]. Now, when a packet is dropped, the correspondent node's TCP implementation is likely to react as if there were a source of network congestion, and initiate the slow-

RED HAT 1010
RED HAT V. CAS

start mechanisms [9] designed for controlling that problem. However, those mechanisms are inappropriate for overcoming errors introduced by the links themselves, and have the effect of magnifying the discontinuity introduced by the dropped packet. This problem has been analyzed by Caceres, et al. [3]; there is no easy solution available, and certainly no solution likely to be installed soon on all correspondent nodes. While this problem is beyond the scope of this document, it does illustrate that providing performance transparency to mobile nodes involves understanding mechanisms outside the network layer. It also indicates the need to avoid designs which systematically drop packets; such designs might otherwise be considered favorably when making engineering tradeoffs.

D. Example Scenarios

This section shows example Registration Requests for several common scenarios.

D.1. Registering with a Foreign Agent Care-of Address

The mobile node receives an Agent Advertisement from a foreign agent and wishes to register with that agent using the advertised foreign agent care-of address. The mobile node wishes only IP-in-IP encapsulation, does not want broadcasts, and does not want simultaneous mobility bindings:

```
IP fields:
  Source Address = mobile node's home address
  Destination Address = copied from the IP source address of the
    Agent Advertisement
  Time to Live = 1
UDP fields:
  Source Port = <any>
  Destination Port = 434
Registration Request fields:
  Type = 1
  S=0,B=0,D=0,M=0,G=0
  Lifetime = the Registration Lifetime copied from the
    Mobility Agent Advertisement Extension of the
    Router Advertisement message
  Home Address = the mobile node's home address
  Home Agent = IP address of mobile node's home agent
  Care-of Address = the Care-of Address copied from the
    Mobility Agent Advertisement Extension of the
    Router Advertisement message
  Identification = Network Time Protocol timestamp or Nonce
Extensions:
  The Mobile-Home Authentication Extension
```


RED HAT 1010
RED HAT V. CAS

D.2. Registering with a Co-Located Care-of Address

The mobile node enters a foreign network that contains no foreign agents. The mobile node obtains an address from a DHCP server [6] for use as a co-located care-of address. The mobile node supports all forms of encapsulation (IP-in-IP, minimal encapsulation, and GRE), desires a copy of broadcast datagrams on the home network, and does not want simultaneous mobility bindings:

```
IP fields:
  Source Address = care-of address obtained from DHCP server
  Destination Address = IP address of home agent
  Time to Live = 64
UDP fields:
  Source Port = <any>
  Destination Port = 434
Registration Request fields:
  Type = 1
  S=0,B=1,D=1,M=1,G=1
  Lifetime = 1800 (seconds)
  Home Address = the mobile node's home address
  Home Agent = IP address of mobile node's home agent
  Care-of Address = care-of address obtained from DHCP server
  Identification = Network Time Protocol timestamp or Nonce
Extensions:
  The Mobile-Home Authentication Extension
```

D.3. Deregistration

The mobile node returns home and wishes to deregister all care-of addresses with its home agent.

IP fields:
  Source Address = mobile node's home address
  Destination Address = IP address of home agent
  Time to Live = 1
UDP fields:
  Source Port = <any>
  Destination Port = 434
Registration Request fields:
  Type = 1
  S=0,B=0,D=0,M=0,G=0
  Lifetime = 0
  Home Address = the mobile node's home address
  Home Agent = IP address of mobile node's home agent
  Care-of Address = the mobile node's home address
  Identification = Network Time Protocol timestamp or Nonce
Extensions:
  The Mobile-Home Authentication Extension

E. Applicability of Prefix Lengths Extension

Caution is indicated with the use of the Prefix Lengths Extension over wireless links, due to the irregular coverage areas provided by wireless transmitters. As a result, it is possible that two foreign agents advertising the same prefix might indeed provide different connectivity to prospective mobile nodes. The Prefix-Lengths Extension SHOULD NOT be included in the advertisements sent by agents in such a configuration.

Foreign agents using different wireless interfaces would have to cooperate using special protocols to provide identical coverage in space, and thus be able to claim to have wireless interfaces situated on the same subnetwork. In the case of wired interfaces, a mobile node disconnecting and subsequently connecting to a new point of attachment, may well send in a new Registration Request no matter whether the new advertisement is on the same medium as the last recorded advertisement. And, finally, in areas with dense populations of foreign agents it would seem unwise to require the propagation via routing protocols of the subnet prefixes associated with each individual wireless foreign agent; such a strategy could lead to quick depletion of available space for routing tables, unwarranted increases in the time required for processing routing updates, and longer decision times for route selection if routes (which are almost always unnecessary) are stored for wireless "subnets".

References

[1] Atkinson, R., "IP Authentication Header", RFC 1826, August 1995.

[2] S. M. Bellovin. Security Problems in the TCP/IP Protocol Suite. ACM Computer Communications Review, 19(2), March 1989.

[3] Ramon Caceres and Liviu Iftode. Improving the Performance of Reliable Transport Protocols in Mobile Computing Environments. IEEE Journal on Selected Areas in Communications, 13(5):850--857, June 1995.

[4] Deering, S., Editor, "ICMP Router Discovery Messages", RFC 1256, September 1991.

[5] Deering, S., "Host Extensions for IP Multicasting", STD 5, RFC 1112, August 1989.

[6] Droms, R., "Dynamic Host Configuration Protocol", RFC 1541, October 1993.

[7] Eastlake, D., Crocker, S., and J. Schiller, "Randomness Requirements for Security", RFC 1750, December 1994.

[8] Hanks, S., Li, R., Farinacci, D., and P. Traina, "Generic Routing Encapsulation (GRE)", RFC 1701, October 1994.

[9] Van Jacobson. Congestion Avoidance and Control. In Proceedings of the SIGCOMM '88 Symposium: Communications Architectures & Protocols, pages 314--329, August 1988.

RED HAT 1010
RED HAT V. CAS

[10] Jacobson, V., "Compressing TCP/IP Headers for Low-Speed Serial Links", RFC 1144, February 1990.

[11] McCloghrie, K., and F. Kastenholz, "Evolution of the Interfaces Group of MIB-II", RFC 1573, January 1994.

[12] McGregor, G., "The PPP Internet Protocol Control Protocol (IPCP)", RFC 1332, May 1992.

[13] Mills, D., "Network Time Protocol (Version 3): Specification, Implementation and Analysis", RFC 1305, March 1992.

[14] Perkins, C., "IP Encapsulation within IP", RFC 2003, October 1996.

[15] Perkins, C., "Minimal Encapsulation within IP", RFC 2004, October 1996.

[16] Plummer, D., "An Ethernet Address Resolution Protocol: Or Converting Network Protocol Addresses to 48.bit Ethernet Addresses for Transmission on Ethernet Hardware", STD 37, RFC 826, November 1982.

[17] Postel, J., "User Datagram Protocol", STD 6, RFC 768, August 1980.

[18] Postel, J., "Multi-LAN Address Resolution", RFC 925, October 1984.

[19] Postel, J., Editor, "Internet Protocol", STD 5, RFC 791, September 1981.

[20] Reynolds, J., and J. Postel, "Assigned Numbers", STD 2, RFC 1700, October 1994.

[21] Rivest, R., "The MD5 Message-Digest Algorithm", RFC 1321, April 1992.

[22] Simpson, W., Editor, "The Point-to-Point Protocol (PPP)", STD 51, RFC 1661, July 1994.

[23] W. Richard Stevens. TCP/IP Illustrated, Volume 1: The Protocols. Addison-Wesley, Reading, Massachusetts, 1994.

RED HAT 1010
RED HAT V. CAS

Editor's Address

Questions about this memo can also be directed to the editor:

Charles Perkins
Room H3-D34
T. J. Watson Research Center
IBM Corporation
30 Saw Mill River Rd.
Hawthorne, NY 10532

Work: +1-914-784-7350
Fax: +1-914-784-6205
EMail: perk@watson.ibm.com

The working group can be contacted via the current chair:

Jim Solomon
Motorola, Inc.
1301 E. Algonquin Rd.
Schaumburg, IL 60196

Work: +1-847-576-2753
EMail: solomon@comm.mot.com

Network Working Group
Request for Comments: 3012
Category: Standards Track

C. Perkins
Nokia Research Center
P. Calhoun
Sun Microsystems Laboratories

November 2000

Mobile IPv4 Challenge/Response Extensions

Status of this Memo

This document specifies an Internet standards track protocol for the Internet community, and requests discussion and suggestions for improvements. Please refer to the current edition of the "Internet Official Protocol Standards" (STD 1) for the standardization state and status of this protocol. Distribution of this memo is unlimited.

Copyright Notice

Copyright (C) The Internet Society (2000). All Rights Reserved.

Abstract

Mobile IP, as originally specified, defines an authentication extension (the Mobile-Foreign Authentication extension) by which a mobile node can authenticate itself to a foreign agent. Unfortunately, this extension does not provide ironclad replay protection for the foreign agent, and does not allow for the use of existing techniques (such as CHAP) for authenticating portable computer devices. In this specification, we define extensions for the Mobile IP Agent Advertisements and the Registration Request that allow a foreign agent to use a challenge/response mechanism to authenticate the mobile node.

RED HAT 1010
RED HAT V. CAS

Table of Contents

1. Introduction 2
2. Mobile IP Agent Advertisement Challenge Extension 3
3. Operation 3
3.1. Mobile Node Processing for Registration Requests . . . 3
3.2. Foreign Agent Processing for Registration Requests . . 5
3.3. Foreign Agent Processing for Registration Replies . . 7
3.4. Home Agent Processing for the Challenge Extensions . . 7
4. MN-FA Challenge Extension 7
5. Generalized Mobile IP Authentication Extension 8
6. MN-AAA Authentication subtype. 9
7. Reserved SPIs for Mobile IP. 9
8. SPI For RADIUS AAA Servers 10
9. Configurable Parameters. 10
10. Error Values 10
11. IANA Considerations 11
12. Security Considerations 12
13. Acknowledgements 12
References 13
A. Verification Infrastructure 14
Addresses 15
Full Copyright Statement 17

1. Introduction

Mobile IP, as originally specified, defines an authentication extension (the Mobile-Foreign Authentication extension) by which a mobile node can authenticate itself to a foreign agent.

Unfortunately, this extension does not provide ironclad replay protection, from the point of view of the foreign agent, and does not allow for the use of existing techniques (such as CHAP [12]) for authenticating portable computer devices. In this specification, we define extensions for the Mobile IP Agent Advertisements and the Registration Request that allow a foreign agent to a use challenge/response mechanism to authenticate the mobile node.

All SPI values defined in this document refer to values for the Security Parameter Index, as defined in RFC 2002 [8]. The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [1].

RED HAT 1010
RED HAT V. CAS

2. Mobile IP Agent Advertisement Challenge Extension

This section defines a new extension to the Router Discovery Protocol [3] for use by foreign agents that need to issue a challenge for authenticating mobile nodes.

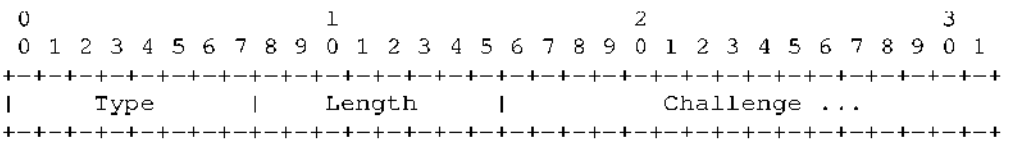


Figure 1: The Challenge Extension

Type 24

Length The length of the Challenge value in bytes; SHOULD be at least 4

Challenge A random value that SHOULD be at least 32 bits.

The Challenge extension, illustrated in figure 1, is inserted in the Agent Advertisements by the Foreign Agent, in order to communicate the latest challenge value that can be used by the mobile node to compute an authentication for its registration request message. The challenge is selected by the foreign agent to provide local assurance that the mobile node is not replaying any earlier registration request. Eastlake, et al. [4] provides more information on generating pseudo-random numbers suitable for use as values for the challenge.

3. Operation

This section describes modifications to the Mobile IP registration process which may occur after the Foreign Agent issues a Mobile IP Agent Advertisement containing the Challenge on its local link.

3.1. Mobile Node Processing for Registration Requests

Whenever the Agent Advertisement contains the Challenge extension, if the mobile node does not have a security association with the Foreign Agent, then it MUST include the Challenge value in a MN-FA Challenge extension to the Registration Request message. If, on the other hand, the mobile node does have a security association with the foreign agent, it SHOULD include the Challenge value in its Registration Request message.

RED HAT 1010
RED HAT V. CAS

If the Mobile Node has a security association with the Foreign Agent, it MUST include a Mobile-Foreign Authentication extension in its Registration Request message, according to the base Mobile IP specification [8]. When the Registration Request contains the MN-FA Challenge extension specified in section 4, the Mobile-Foreign Authentication MUST follow the Challenge extension in the Registration Request.

If the Mobile Node does not have a security association with the Foreign Agent, the Mobile Node MUST include the MN-AAA Authentication extension as defined in section 6. In addition, the Mobile Node SHOULD include the NAI extension [2], to enable the foreign agent to make use of any available verification infrastructure. The SPI field of the MN-AAA Authentication extension specifies the particular secret and algorithm (shared between the Mobile Node and the verification infrastructure) that must be used to perform the authentication. If the SPI value is chosen as CHAP_SPI (see section 9), then the mobile node specifies CHAP-style authentication [12] using MD5 [11].

In either case, the MN-FA Challenge extension and one of the above specified authentication extensions MUST follow the Mobile-Home Authentication extension, if present.

A successful Registration Reply from the Foreign Agent MAY include a new Challenge value (see section 3.3). The Mobile Node MAY use either the value found in the latest Advertisement, or the one found in the last Registration Reply from the Foreign Agent. This approach enables the Mobile Node to make use of the challenge without having to wait for advertisements.

A Mobile Node might receive an UNKNOWN_CHALLENGE error (see section 9) if it moves to a new Foreign Agent that cannot validate the challenge provided in the Registration Request. In such instances, the Mobile Node MUST use a new Challenge value in any new registration, obtained either from an Agent Advertisement, or from a Challenge extension to the Registration Reply containing the error.

A Mobile Node that does not include a Challenge when the Mobile-Foreign Authentication extension is present may receive a MISSING_CHALLENGE (see section 10) error. In this case, the foreign agent will not process the request from the mobile node unless the request contains a valid Challenge.

RED HAT 1010
RED HAT V. CAS

A Mobile Node that receives a BAD_AUTHENTICATION error code (see section 10) SHOULD include the MN-AAA Authentication Extension in the next Registration Request. This will make it possible for the Foreign Agent to use its AAA infrastructure in order to authenticate the Mobile Node.

3.2. Foreign Agent Processing for Registration Requests

Upon receipt of the Registration Request, if the Foreign Agent has issued a Challenge as part of its Agent Advertisements, and it does not have a security association with the mobile node, then the Foreign Agent MUST check that the MN-FA Challenge extension exists, and that it contains a challenge value previously unused by the Mobile Node. This ensures that the mobile node is not attempting to replay a previous advertisement and authentication. If the challenge extension is needed and does not exist, the Foreign Agent MUST send a Registration Reply to the mobile node with the error code MISSING_CHALLENGE.

A foreign agent that sends Agent Advertisements containing a Challenge value MAY send a Registration Reply message with a MISSING_CHALLENGE error if the mobile node sends a Registration Request with a Mobile-Foreign Authentication extension without including a Challenge. In other words, such a foreign agent MAY refuse to process a Registration Request request from the mobile node unless the request contains a valid Challenge.

If a mobile node retransmits a Registration Request with the same Identification field and the same Challenge extension, and the Foreign Agent still has a pending Registration Request record in effect for the mobile node, then the Foreign Agent forwards the Registration Request to the Home Agent again. In all other circumstances, if the Foreign Agent receives a Registration Request with a Challenge extension containing a Challenge value previously used by that mobile node, the Foreign Agent SHOULD send a Registration Reply to the mobile node containing the Code value STALE_CHALLENGE.

The Foreign Agent MUST NOT accept any Challenge in the Registration Request unless it was offered in last successful Registration Reply issued to the Mobile Node, or else advertised as one of the last CHALLENGE_WINDOW (see section 9) Challenge values inserted into the immediately preceding Agent advertisements. If the Challenge is not one of the recently advertised values, the foreign Agent SHOULD send a Registration Reply with Code UNKNOWN_CHALLENGE (see section 10).

RED HAT 1010
RED HAT V. CAS

Furthermore, the Foreign Agent MUST check that there is either a Mobile-Foreign, or a MN-AAA Authentication extension after the Challenge extension. Any registration message containing the Challenge extension without either of these authentication extensions MUST be silently discarded. If the registration message contains a Mobile-Foreign Authentication extension with an incorrect authenticator that fails verification, the Foreign Agent MAY send a Registration Reply to the mobile node with Code value BAD_AUTHENTICATION (see Section 10).

If the MN-AAA Authentication extension (see Section 6) is present in the message, or if an NAI extension is included indicating that the mobile node belongs to a different administrative domain, the foreign agent may take actions outside the scope of this protocol specification to carry out the authentication of the mobile node. The Foreign Agent MUST NOT remove the MN-AAA Authentication Extension from the Registration Request prior to the completion of the authentication performed by the AAA infrastructure. The appendix provides an example of an action that could be taken by a foreign agent.

In the event that the Challenge extension is authenticated through the Mobile-Foreign Authentication Extension, the Foreign Agent MAY remove the Challenge Extension from the Registration Request without disturbing the authentication value computed by the Mobile Node for use by the AAA or the Home Agent. If the Challenge extension is not removed, it MUST precede the Foreign-Home Authentication extension.

If the Foreign Agent does not remove the Challenge extension, then the Foreign Agent SHOULD store the Challenge value as part of the pending registration request list [8]. Also in this case, the Foreign Agent MUST reject any Registration Reply message coming from the Home Agent that does not also include the Challenge Extension with the same Challenge Value that was included in the Registration Request. The Foreign Agent MUST send the rejected Registration message to the mobile node, and change the status in the Registration Reply to the value MISSING_CHALLENGE (see section 10).

If the Foreign Agent does remove the Challenge extension and applicable authentication from the Registration Request message, then it SHOULD insert the Identification field from the Registration Request message along with its record-keeping information about the particular Mobile Node in order to protect against replays.

RED HAT 1010
RED HAT V. CAS

3.3. Foreign Agent Processing for Registration Replies

The Foreign Agent MAY include a new Challenge extension in any Registration Reply, successful or not. If the foreign agent includes this extension in a successful Registration Reply, the extension SHOULD precede a MN-FA authentication extension.

Suppose the Registration Reply includes a Challenge extension from the Home Agent, and the foreign agent wishes to include another Challenge extension with the Registration Reply for use by the mobile node. In that case, the foreign agent MUST delete the Challenge extension from the Home Agent from the Registration Reply, along with any FA-HA authentication extension, before appending the new Challenge extension to the Registration Reply.

3.4. Home Agent Processing for the Challenge Extensions

If the Home Agent receives a Registration Request with the MN-FA Challenge extension, and recognizes the extension, the Home Agent MUST include the Challenge extension in the Registration Reply. The Challenge Extension MUST be placed after the Mobile-Home authentication extension, and the extension SHOULD be authenticated by a Foreign-Home Authentication extension.

Since the extension type for the Challenge extension is within the range 128-255, the Home Agent MUST process such a Registration Request even if it does not recognize the Challenge extension [8]. In this case, the Home Agent will send a Registration Reply to the Foreign Agent that does not include the Challenge extension.

4. MN-FA Challenge Extension

This section specifies a new Mobile IP Registration extension that is used to satisfy a Challenge in an Agent Advertisement. The Challenge extension to the Registration Request message is used to indicate the challenge that the mobile node is attempting to satisfy.

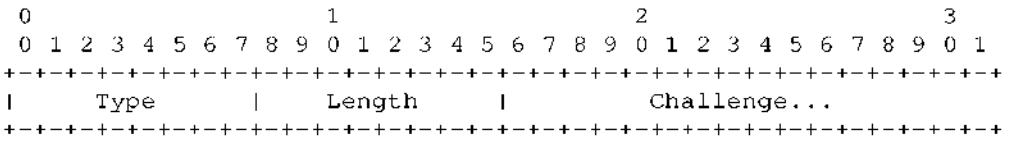


Figure 2: The MN-FA Challenge Extension

Type 132 (skippable) (see [8])

Length Length of the Challenge value

RED HAT 1010
RED HAT V. CAS

Challenge The Challenge field is copied from the Challenge field found in the Agent Advertisement Challenge extension (see section 2).

5. Generalized Mobile IP Authentication Extension

Several new authentication extensions have been designed for various control messages proposed for extensions to Mobile IP (see, for example, [9]). A new authentication extension is required for a mobile node to present its credentials to any other entity other than the ones already defined; the only entities defined in the base Mobile IP specification [8] are the home agent and the foreign agent. It is the purpose of the generalized authentication extension defined here to collect together data for all such new authentication applications into a single extension type with subtypes.

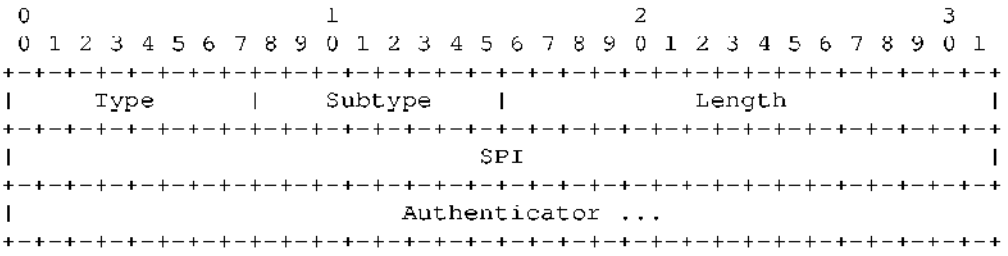


Figure 3: The Generalized Mobile IP Authentication Extension

Type 36 (not skippable) (see [8])

Subtype a number assigned to identify the kind of endpoints or characteristics of the particular authentication strategy

Length 4 plus the number of bytes in the Authenticator; MUST be at least 20.

SPI Security Parameters Index

Authenticator The variable length Authenticator field

In this document, only one subtype is defined:

1 MN-AAA Authentication subtype (see section 6)

RED HAT 1010
RED HAT V. CAS

6. MN-AAA Authentication subtype

The Generalized Authentication extension with subtype 1 will be referred to as a MN-AAA Authentication extension. If the mobile node does not include a Mobile-Foreign Authentication [8] extension, then it MUST include the MN-AAA Authentication extension whenever the Challenge extension is present. If the MN-AAA Authentication extension is present, then the Registration Message sent by the mobile node MUST contain the Mobile-HA Authentication extension [8] if it shares a security association with the Home Agent. If present, the Mobile-HA Authentication Extension MUST appear prior to the MN-AAA Authentication extension. The mobile node MAY include a MN-AAA Authentication extension in any Registration Request. The corresponding response MUST include the MN-HA Authentication Extension, and MUST NOT include the MN-AAA Authentication Extension.

The default algorithm for computation of the authenticator is HMAC-MD5 [5] computed on the following data, in the order shown:

```
Preceding Mobile IP data || Type, Subtype, Length, SPI
```

where the Type, Length, Subtype, and SPI are as shown in section 5. The resulting function call, as described in [5], would be:

```
hmac_md5(data, datalen, Key, KeyLength, authenticator);
```

Each mobile node MUST support the ability to produce the authenticator by using HMAC-MD5 as shown. Just as with Mobile IP, this default algorithm MUST be able to be configured for selection at any arbitrary 32-bit SPI outside of the SPIs in the reserved range 0-255.

7. Reserved SPIs for Mobile IP

Mobile IP defines several authentication extensions for use in Registration Requests and Replies. Each authentication extension carries a Security Parameters Index (SPI) which should be used to index a table of security associations. Values in the range 0 - 255 are reserved for special use. A list of reserved SPI numbers is to be maintained by IANA at the following URL:

```
http://www.iana.org/numbers.html
```


RED HAT 1010
RED HAT V. CAS

8. SPI For RADIUS AAA Servers

Some AAA servers only admit a single security association, and thus do not use the SPI numbers for Mobile IP authentication extensions for use when determining the security association that would be necessary for verifying the authentication information included with the Authentication extension.

SPI number CHAP_SPI (see section 9) is reserved (see section 7) for indicating the following procedure for computing authentication data (called the "authenticator"), which is used by many RADIUS servers [10] today.

To compute the authenticator, apply MD5 [11] computed on the following data, in the order shown:

```
High-order byte from Challenge || Key ||
MD5(Preceding Mobile IP data ||
Type, Subtype (if present), Length, SPI) ||
Least-order 237 bytes from Challenge
```

where the Type, Length, SPI, and possibly Subtype, are the fields of the authentication extension in use. For instance, all four of these fields would be in use when SPI == CHAP_SPI is used with the Generalized Authentication extension. Since the RADIUS protocol cannot carry attributes greater than 253 in size, the preceding Mobile IP data, type, subtype (if present), length and SPI are hashed using MD5. Finally, the least significant 237 bytes of the challenge are concatenated.

9. Configurable Parameters

Every Mobile IP agent supporting the extensions defined in this document SHOULD be able to configure each parameter in the following table. Each table entry contains the name of the parameter, the default value, and the section of the document in which the parameter first appears.

Parameter Name	Default Value	Section(s) of Document
CHALLENGE_WINDOW	2	3.2
CHAP_SPI	2	8

10. Error Values

Each entry in the following table contains the name of Code [8] to be returned in a Registration Reply, the value for the Code, and the section in which the error is first mentioned in this specification.

RED HAT 1010
RED HAT V. CAS

Error Name	Value	Section of Document
UNKNOWN_CHALLENGE	104	3.2
BAD_AUTHENTICATION	67	3.2 - also see [8]
MISSING_CHALLENGE	105	3.1,3.2
STALE_CHALLENGE	106	3.2

11. IANA Considerations

The Generalized Mobile IP Authentication extension defined in Section 5 is a Mobile IP registration extension as defined in RFC 2002 [8] and extended in RFC 2356 [7]. IANA should assign a value of 36 for this extension.

A new number space is to be created for enumerating subtypes of the Generalized Authentication extension (see section 5). New subtypes of the Generalized Authentication extension, other than the number (1) for the MN-AAA authentication extension specified in section 6, must be specified and approved by a designated expert.

The MN-FA Challenge Extension defined in Section 4 is a router advertisement extension as defined in RFC 1256 [3] and extended in RFC 2002 [8]. IANA should assign a value of 132 for this purpose.

The Code values defined in Section 10 are error codes as defined in RFC 2002 [8] and extended in RFC 2344 [6] and RFC 2356 [7]. They correspond to error values conventionally associated with rejection by the foreign agent (i.e., values from the range 64-127). The Code value 67 is a pre-existing value which is to be used in some cases with the extension defined in this specification. IANA should record the values as defined in Section 10.

A new section for enumerating algorithms identified by specific SPIs within the range 0-255 is to be added to

http://www.isi.edu/in-notes/iana/assignments/mobileip-numbers.

The CHAP_SPI number (2) discussed in section 8 is to be assigned from this range of reserved SPI numbers. New assignments from this reserved range must be specified and approved by the Mobile IP working group. SPI number 1 should not be assigned unless in the future the Mobile IP working group decides that SKIP is not important for enumeration in the list of reserved numbers. SPI number 0 should not be assigned.

RED HAT 1010
RED HAT V. CAS

12. Security Considerations

In the event that a malicious mobile node attempts to replay the authenticator for an old MN-FA Challenge, the Foreign Agent would detect it since the agent always checks whether it has recently advertised the Challenge (see section 3.2). Allowing mobile nodes with different IP addresses or NAIs to use the same Challenge value does not represent a security vulnerability, because the authentication data provided by the mobile node will be computed over data that is different (at least by the bytes of the mobile nodes' IP addresses).

Whenever a Foreign Agent updates a field of the Registration Reply (as suggested in section 3.2), it invalidates the authentication data supplied by the Home Agent in the MN-HA Authentication extension to the Registration Reply. Thus, this opens up a security exposure whereby a node might try to supply a bogus Registration Reply to a mobile node that causes the mobile node to act as if its Registration Reply were rejected. This might happen when, in fact, a Registration Reply showing acceptance of the registration might soon be received by the mobile node.

If the foreign agent chooses a Challenge value (see section 2) with fewer than 4 bytes, the foreign agent SHOULD maintain records that also the Identification field for the mobile node. The foreign agent can then find assurance that the Registration messages using the short Challenge value are in fact unique, and thus assuredly not replayed from any earlier registration.

Section 8 (SPI For RADIUS AAA Servers) defines a method of computing the Generalized Mobile IP Authentication Extension's authenticator field using MD5 in a manner that is consistent with RADIUS [10]. The use of MD5 in the method described in Section 8 is less secure than HMAC-MD5 [5], and should be avoided whenever possible.

13. Acknowledgements

The authors would like to thank Tom Hiller, Mark Munson, the TIA TR45-6 WG, Gabriel Montenegro, Vipul Gupta, and Pete McCann for their useful discussions. A recent draft by Mohamed Khalil, Raja Narayanan, Emad Qaddoura, and Haseeb Akhtar has also suggested the definition of a generalized authentication extension similar to the specification contained in section 5.

RED HAT 1010
RED HAT V. CAS

References

[1] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[2] Calhoun, P. and C. Perkins. "Mobile IP Network Access Identifier Extension for IPv4", RFC 2794, January 2000.

[3] Deering, S., "ICMP Router Discovery Messages", RFC 1256, September 1991.

[4] Eastlake, D., Crocker, S. and J. Schiller, "Randomness Recommendations for Security", RFC 1750, December 1994.

[5] Krawczyk, H., Bellare, M. and R. Canetti, "HMAC: Keyed-Hashing for Message Authentication", RFC 2104, February 1997.

[6] Montenegro, G., "Reverse Tunneling for Mobile IP", RFC 2344, May 1998.

[7] Montenegro, G. and V. Gupta, "Sun's SKIP Firewall Traversal for Mobile IP", RFC 2356, June 1998.

[8] Perkins, C., "IP Mobility Support", RFC 2002, October 1996.

[9] Perkins, C. and D. Johnson, "Route Optimization in Mobile IP", Work in Progress.

[10] Rigney, C., Rubens, A., Simpson, W. and S. Willens, "Remote Authentication Dial In User Service (RADIUS)", RFC 2138, April 1997.

[11] Rivest, R., "The MD5 Message-Digest Algorithm", RFC 1321, April 1992.

[12] Simpson, W., "PPP Challenge Handshake Authentication Protocol (CHAP)", RFC 1994, August 1996.

RED HAT 1010
RED HAT V. CAS

A. Verification Infrastructure

The Challenge extensions in this protocol specification are expected to be useful to help the Foreign Agent manage connectivity for visiting mobile nodes, even in situations where the foreign agent does not have any security association with the mobile node or the mobile node's home agent. In order to carry out the necessary authentication, it is expected that the foreign agent will need the assistance of external administrative systems, which have come to be called AAA systems. For the purposes of this document, we call the external administrative support the "verification infrastructure". The verification infrastructure is described to motivate the design of the protocol elements defined in this document, and is not strictly needed for the protocol to work. The foreign agent is free to use any means at its disposal to verify the credentials of the mobile node. This could, for instance, rely on a separate protocol between the foreign agent and the Mobile IP home agent, and still be completely invisible to the mobile node.

In order to verify the credentials of the mobile node, we imagine that the foreign agent has access to a verification infrastructure that can return a secure notification to the foreign agent that the authentication has been performed, along with the results of that authentication. This infrastructure may be visualized as shown in figure 4.

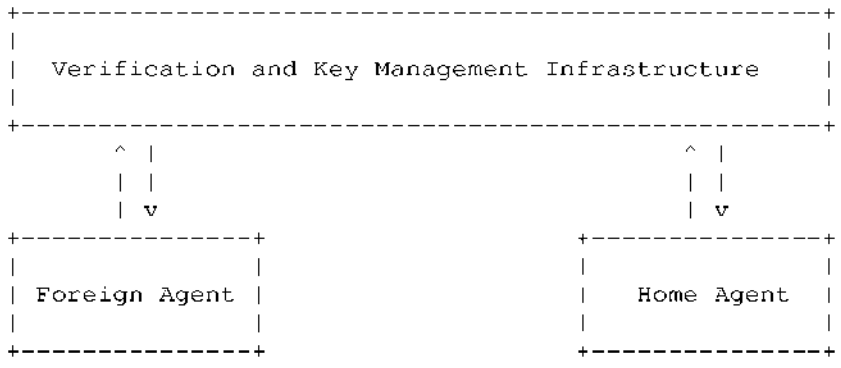


Figure 4: The Verification Infrastructure

After the foreign agent gets the Challenge authentication, it MAY pass the authentication to the (here unspecified) infrastructure, and await a Registration Reply. If the Reply has a positive status (indicating that the registration was accepted), the foreign agent

RED HAT 1010
RED HAT V. CAS

accepts the registration. If the Reply contains the Code value BAD_AUTHENTICATION (see Section 10), the foreign agent takes actions indicated for rejected registrations.

Implicit in this picture, is the important observation that the Foreign Agent and the Home Agent have to be equipped to make use of whatever protocol is made available to them by the challenge verification and key management infrastructure shown in the figure.

The protocol messages for handling the authentication within the verification infrastructure, and identity of the agent performing the verification of the Foreign Agent challenge, are not specified in this document, because those operations do not have to be performed by any Mobile IP entity.

Addresses

The working group can be contacted via the current chairs:

Basavaraj Patil
Nokia Corporation
6000 Connection Drive
M/S M8-540
Irving, Texas 75039
USA

Phone: +1 972-894-6709
Fax : +1 972-894-5349
EMail: Basavaraj.Patil@nokia.com

Phil Roberts
Motorola
1501 West Shure Drive
Arlington Heights, IL 60004
USA

Phone:+1 847-632-3148
EMail: QA3445@email.mot.com

RED HAT 1010
RED HAT V. CAS

Questions about this memo can also be directed to the authors:

Charles E. Perkins
Communications Systems Lab
Nokia Research Center
313 Fairchild Drive
Mountain View, California 94043
USA

Phone: +1-650 625-2986
Fax: +1 650 625-2502
EMail: charliep@iprg.nokia.com

Pat R. Calhoun
Network & Security Center
Sun Microsystems Laboratories
15 Network Circle
Menlo Park, California 94025
USA

Phone: +1 650-786-7733
Fax: +1 650-786-6445
EMail: pcalhoun@eng.sun.com

RED HAT 1010
RED HAT V. CAS

Full Copyright Statement

Copyright (C) The Internet Society (2000). All Rights Reserved.

This document and translations of it may be copied and furnished to others, and derivative works that comment on or otherwise explain it or assist in its implementation may be prepared, copied, published and distributed, in whole or in part, without restriction of any kind, provided that the above copyright notice and this paragraph are included on all such copies and derivative works. However, this document itself may not be modified in any way, such as by removing the copyright notice or references to the Internet Society or other Internet organizations, except as needed for the purpose of developing Internet standards in which case the procedures for copyrights defined in the Internet Standards process must be followed, or as required to translate it into languages other than English.

The limited permissions granted above are perpetual and will not be revoked by the Internet Society or its successors or assigns.

This document and the information contained herein is provided on an "AS IS" basis and THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIMS ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Acknowledgement

Funding for the RFC Editor function is currently provided by the Internet Society.

RED HAT 1010
RED HAT V. CAS

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION NO.:	**14/512,414**	EXAMINER:	
INVENTOR:	Eric M. Delangis	GROUP ART UNIT:	2652
FILED:	OCTOBER 11, 2014	CONFIRMATION NO.:	9530
TITLE:	Multipath Communication Devices and Methods		
ATTORNEY NO:	ED 2000		

Commissioner For Patents
PO Box 1450
Alexandria, VA 22313-1450

PRELIMINARY AMENDMENT

This preliminary amendment is being submitted before an office action has been issued. No additional fees are required because no new claims are being submitted. All the pending claims are being amended. It is respectfully submitted that the amended claims do not contain any new matter.

CERTIFICATE OF ELECTRONIC (EFS-WEB) TRANSMISSION

I hereby certify that this correspondence is being transmitted via the U.S. Patent Office's electronic filing system in accordance with 37 C.F.R. § 1.8(a)(1) from the **Pacific Time Zone** of the United States on the local date shown below.

Dated: July 23, 2015

By: Clay McGurk
Clay McGurk

RED HAT 1010
RED HAT V. CAS

CLAIMS

This listing of claims will replace all prior versions and listings of the claims in the application:

1. (Currently Amended) A method for establishing and sending data across multiple paths between a ~~server~~ host and an initiating device, the host and a plurality of devices capable of being connected to one or more networks, comprising:

receiving at the ~~server~~ host information about ~~of~~ each of a plurality of devices that are capable of joining ~~selected to participate in~~ a multipath connection with the initiating device~~, the multipath connection comprising each the devices being connected with the initiating device~~; and

the ~~server~~ host performing at least one of (i) ~~dividing data into groups of data packets and~~ sending ~~different ones of the~~ groups of data packets to ~~each~~ one or more of the devices that are participating in the multipath connection, (ii) sending separate streams of data to ~~each~~ one or more of the devices that are participating in the multipath connection, and (iii) sending combinations of separate streams and groups of data packets to ~~each~~ one or more of the devices that are participating in the multipath connection.

2. (Currently Amended) The method of claim 1, wherein the information includes a connection ID that is [[is]] generated by one of the ~~server~~ host, one of the plurality of devices, the initiating device and a network device.

3. (Currently Amended) The method of claim 2, further comprising receiving at the ~~server~~ host information ~~of~~ about one or more additional devices ~~able to participate in~~ capable of joining the multipath connection, the one or more additional devices being associated with the connection ID.

4. (Currently Amended) The method of claim 2, further comprising receiving at the ~~server~~ host information about which devices are to be removed from the multipath connection, the devices to be removed being associated with the connection ID.

5. (Currently Amended) The method of claim 1, further comprising at least one of the ~~server~~ host, the devices and the initiating device determining which of the devices are to be removed from the multipath connection.

6. (Currently Amended) The method of claim 1, wherein the ~~server~~ host is one of a multipath link aggregator, a network relay, ~~and~~ a multipath link server, a server, a network switch, a network router and a device capable of providing multipath services.

7. (Currently Amended) A method for establishing and sending data across multiple paths in different multipath connections from a ~~server~~ host, comprising:

the host ~~server~~ receiving ~~connection~~ multipath requests, each of the multipath requests involving a different ~~for a plurality of~~ multipath ~~connections~~ connection;

the host ~~server~~ associating a connection ID for each ~~connection~~ multipath request ~~associated with each of the multipath connections~~;

the host ~~server~~ receiving information of each of a plurality of devices capable of joining ~~participating~~ in a particular multipath connection having a particular connection ID; and

the host ~~server~~ sending data associated with the particular connection ID to ~~each~~ one or more of the devices participating in the particular multipath connection associated with the particular connection ID.

8. (Currently Amended) The method of claim 7, further comprising receiving at the ~~server~~ host information ~~of additional~~ about one or more devices ~~participating in~~ capable of joining the particular multipath connection, the ~~additional~~ one or more devices being associated with the particular connection ID.

9. (Currently Amended) The method of claim 7, further comprising receiving at the ~~server~~ host identification of the devices to be removed from the particular multipath connection, the devices to be removed being associated with the particular connection ID.

RED HAT 1010
RED HAT V. CAS

10. (Currently Amended) The method of claim 7, further comprising at least one of the ~~server~~ host and the devices determining which of the devices are to be removed from the particular multipath connection.

11. (Currently Amended) The method of claim 7, wherein the ~~server~~ host is one of a multipath link aggregator, a network relay, ~~and~~ a multipath link server, a server, a network switch, a network router and a device capable of providing multipath services.

12. (Currently Amended) The method of claim 7, wherein the connection ID is generated by one of the ~~server~~ host, one of the plurality of devices, ~~or~~ and a network device.

13. (Currently Amended) A method for establishing and sending data across multiple paths between a device and a ~~server~~ host, the device capable of being connected to at least one of a network and one or more network devices, the remote devices capable of being connected to at least one network, the host capable of being connected to one or more networks, comprising:

identifying ~~selecting~~ one or more ~~remote~~ network devices ~~to participate in~~ capable of joining a multipath connection with the device~~, wherein the multipath connection comprises a connection between the device and each of the selected one or more remote devices~~; and

receiving data from the host ~~concurrently~~ at the device via one or more of the ~~selected one or more remote~~ identified network devices during the multipath connection.

14. (Currently Amended) The method of claim 13, further comprising receiving at the device information about the one or more ~~remote~~ network devices, the information including a confirmation of a capability of ~~participating~~ joining in the multipath connection.

15. (Currently Amended) The method of claim 13, wherein the network device is one of a cell phone, a tablet, a router, a computer, a Wi-Fi device, a Wi-Max device, a communications device, a mobile communications device, and a Bluetooth device.

16. (Currently Amended) The method of claim 13, wherein the network devices ~~are~~ comprise one or more routers, modems, cell phones, tablets, computers, cellular towers, wireless towers, base stations, wireless base stations, repeaters, wireless repeaters, Wi-Fi devices, networking devices, hot spot devices, Wi-MAX devices, LTE devices, GSM devices, consumer electronic devices, ~~and~~ an Internet- ready device, a network switch, a network router, a multipath server and a device capable of providing multipath services.

17. (Currently Amended) The method of claim 13, wherein the selected one or more ~~remote~~ network devices is capable of connecting to another ~~remote~~ network device which is capable of ~~participating in~~ joining the multipath connection, and further comprising relaying data from the another ~~remote~~ network device to the device through the selected one or more ~~remote~~ network devices.

RED HAT 1010
RED HAT V. CAS

18. (Currently Amended) A communications device comprising:

at least one interface capable of connecting concurrently to at least one of a plurality of ~~remote~~ network devices; and

a processor coupled to the at least one interface that

identifies ~~selects~~ one or more ~~remote~~ network devices ~~to participate in~~ capable of joining a multipath connection with the communications device~~, wherein the multipath connection comprises a connection between the device and each of the selected one or more remote devices~~;

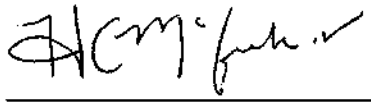
receives ~~receiving~~ data ~~concurrently~~ at the device via one or more of the identified ~~the selected one or more remote~~ network devices during the multipath connection.

19. (Currently Amended) The communications device of claim ~~19~~ 18, wherein the device is one of a cell phone, a tablet, a router, a computer, a Wi-Fi device, a Wi-Max device, a mobile communications device, and a Bluetooth device.

20. (Currently Amended) The communications device of claim ~~19~~ 18, wherein each of the ~~remote~~ network devices comprises one or more communications devices, cellular devices, tablet devices, router devices, computer devices, satellite devices, Wi-MAX devices, wireless devices, mobile communications devices, Bluetooth devices, and Wi-Fi devices, a network switch, a network router, a multipath server and a device capable of providing multipath services.

RED HAT 1010
RED HAT V. CAS

Respectfully Submitted,



Date: July 23, 2015

H C McGurk IV, Reg. No. 34,964
Attorney for Applicant

The Law Office of Clay McGurk
PO Box 1488
Orange, CA 92856
Phone: **949.395.3942**
Email: claymcgurk@gmail.com

Electronic Acknowledgement Receipt

EFS ID:	23009055
Application Number:	14512414
International Application Number:	
Confirmation Number:	9530
Title of Invention:	MULTIPATH COMMUNICATION DEVICES AND METHODS
First Named Inventor/Applicant Name:	Eric M. Delangis
Customer Number:	64828
Filer:	Harold C. McGurk
Filer Authorized By:	
Attorney Docket Number:	ED 2000
Receipt Date:	23-JUL-2015
Filing Date:	11-OCT-2014
Time Stamp:	17:53:28
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Preliminary Amendment	14512414_Preliminary_Amendment_7-23-2015.pdf	97380 2eb092 1f8c03a82211000e7 1old6cf1ae3 14f b685	no	7

Warnings:	
Information:	

RED HAT 1010
RED HAT V. CAS

Total Files Size (in bytes):	97380

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111
If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371
If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office
If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

RED HAT 1010
RED HAT V. CAS

PTO/SB/06 (09-11)
Approved for use through 1/31/2014 OMB 0651-0032
U.S. Patent and Trademark Office, U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PATENT APPLICATION FEE DETERMINATION RECORD Substitute for Form PTO-875	Application or Docket Number 14/512,414	Filing Date 10/11/2014	☒ To be Mailed

ENTITY: ☐ LARGE ☒ SMALL ☐ MICRO

APPLICATION AS FILED – PART I

FOR	(Column 1) NUMBER FILED	(Column 2) NUMBER EXTRA	RATE ($)	FEE ($)
☐ BASIC FEE (37 CFR 1.16(a), (b), or (c))	N/A	N/A	N/A	
☐ SEARCH FEE (37 CFR 1.16(k), (l), or (m))	N/A	N/A	N/A	
☐ EXAMINATION FEE (37 CFR 1.16(o), (p), or (q))	N/A	N/A	N/A	
TOTAL CLAIMS (37 CFR 1.16(i))	minus 20 = *		X $ =	
INDEPENDENT CLAIMS (37 CFR 1.16(h))	minus 3 = *		X $ =	
☐ APPLICATION SIZE FEE (37 CFR 1.16(s))	If the specification and drawings exceed 100 sheets of paper, the application size fee due is $310 ($155 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).			
☐ MULTIPLE DEPENDENT CLAIM PRESENT (37 CFR 1.16(j))				
* If the difference in column 1 is less than zero, enter "0" in column 2.			TOTAL	

APPLICATION AS AMENDED – PART II

AMENDMENT	07/23/2015	(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	RATE ($)	ADDITIONAL FEE ($)
	Total (37 CFR 1.16(i))	* 20	Minus	** 20	= 0	X $40 =	0
	Independent (37 CFR 1.16(h))	* 4	Minus	***4	= 0	X $210 =	0
	☐ Application Size Fee (37 CFR 1.16(s))						
	☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))						
						TOTAL ADD'L FEE	0

AMENDMENT		(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	RATE ($)	ADDITIONAL FEE ($)
	Total (37 CFR 1.16(i))	*	Minus	**	=	X $ =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=	X $ =	
	☐ Application Size Fee (37 CFR 1.16(s))						
	☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))						
						TOTAL ADD'L FEE	

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.
** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20".
*** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3".
The "Highest Number Previously Paid For" (Total or Independent) is the highest number found in the appropriate box in column 1.

LIE
/ANGELA JOHNSON/

This collection of information is required by 37 CFR 1.16. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**
If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

RED HAT 1010
RED HAT V. CAS

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
14/512,414	10/11/2014	Eric M. Delangis	ED 2000	9530

64828 7590 08/18/2015

Harold McGurk
The Law Office of Clay McGurk
P.O. BOX 1488
Orange, CA 92856

EXAMINER	
CHO, HONG SOL	

ART UNIT	PAPER NUMBER
2467	

NOTIFICATION DATE	DELIVERY MODE
08/18/2015	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

CLAYMCGURK@GMAIL.COM

Office Action Summary	Application No. 14/512,414	Applicant(s) DELANGIS, ERIC M.	
	Examiner HONG CHO	Art Unit 2467	AIA (First Inventor to File) Status No

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE *3* MONTHS FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☐ Responsive to communication(s) filed on _____.

☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on _____.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ An election was made by the applicant in response to a restriction requirement set forth during the interview on _____; the restriction requirement and election have been incorporated into this action.

4)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims*

5)☒ Claim(s) *1-20* is/are pending in the application.

5a) Of the above claim(s) _____ is/are withdrawn from consideration.

6)☐ Claim(s) _____ is/are allowed.

7)☒ Claim(s) *1-20* is/are rejected.

8)☐ Claim(s) _____ is/are objected to.

9)☐ Claim(s) _____ are subject to restriction and/or election requirement.

* If any claims have been determined allowable, you may be eligible to benefit from the **Patent Prosecution Highway** program at a participating intellectual property office for the corresponding application. For more information, please see http://www.uspto.gov/patents/init_events/pph/index.jsp or send an inquiry to PPHfeedback@uspto.gov.

Application Papers

10)☐ The specification is objected to by the Examiner.

11)☒ The drawing(s) filed on *10/11/2014* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

Certified copies:

a)☐ All b)☐ Some** c)☐ None of the:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

** See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)

2) ☒ Information Disclosure Statement(s) (PTO/SB/08a and/or PTO/SB/08b) Paper No(s)/Mail Date _____.

3) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .

4) ☐ Other: _____.

1. The present application is being examined under the pre-AIA first to invent provisions.

DETAILED ACTION

Claim Rejections - 35 USC § 102

2. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102(e) that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless

(e) the invention was described in (1) an application for patent, published under section 122(b), by another filed in the United States before the invention by the applicant for patent or (2) a patent granted on an application for patent by another filed in the United States before the invention by the applicant for patent, except that an international application filed under the treaty defined in section 351(a) shall have the effects for purposes of this subsection of an application filed in the United States only if the international application designated the United States and was published under Article 21(2) of such treaty in the English language.

3. Claims 1-3, 6-8 and 11-20 are rejected under 35 U.S.C. 102(e) as being anticipated by Perkins et al (US 6496477), hereinafter referred to as Perkins.

Re claims 1 and 7, Perkins receiving at the host (figure 12, element 1121) information about each of a plurality of devices (nodes in packet network 100) that are capable of joining a multipath connection with the initiating device (figure 12, element 1111) and the host performing at least one of (i) dividing data into groups of data packets and sending different ones of the groups of data packets to each one or more of the

RED HAT 1010
RED HAT V. CAS

devices that are participating in the multipath connection, (ii) sending separate streams of data to each one or more of the devices that are participating in the multipath connection, and (iii) sending combinations of separate streams and groups of data packets to one or more of the devices that are participating in the multipath connection (column 24, lines 26-48).

Re claims 2, 3, 7, 8, 12 and 14, Perkins discloses the host associating a connection ID for each connection multipath request associated with each of the multipath connections and the host sending data associated with the particular connection ID to one or more of the devices participating in the particular multipath connection associated with the particular connection ID (figure 12; multicasts communication stream based on group membership, column 24, lines 18-25).

Re claims 6, 11, 15, 16, 19 and 20, Perkins discloses the device is one of a multipath link aggregator, a network relay, and a multipath link server, a server, a network switch, a network router and a device capable of providing multipath services (figure 12, elements 1111, 1121 and 1125).

Re claims 13 and 18, Perkins discloses identifying one or more remote network devices to participate in capable of joining a multipath connection with the device, and receiving data from the host at the device via one or more of the identified network devices during the multipath connection (figure 12; column 24, lines 26-48).

Re claim 17, Perkins inherently discloses relaying data from another remote network device to the device through the selected one or more remote network devices (figure 12).

RED HAT 1010
RED HAT V. CAS

Claim Rejections - 35 USC § 103

4. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

5. Claims 4, 5, 9 and 10 are rejected under 35 U.S.C. 103(a) as being unpatentable over Perkins in view of Pung et al (US 20020150099), hereinafter referred to as Pung.

Re claims 4, 5, 9 and 10, Perkins discloses all of the limitations of the base claim, but fails to disclose receiving at the host information about which devices are to be removed from the multipath connection, the devices to be removed being associated with the connection ID. However, Pung discloses receiving a prune-branch packet from a joined computing device and removing the branch from the multicast tree (paragraph [0079]; figure 8a). Therefore, it would have been obvious to one having ordinary skill in the art at the time the invention was made to modify the system of Perkins with Pung for the benefit of providing efficient utilization of network resources by removing an entry in multicast tree table.

Conclusion

RED HAT 1010
RED HAT V. CAS

6. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Hong Cho whose telephone number is 571-272-3087. The examiner can normally be reached on Mon-Fri during 7 am to 4 pm.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Hassan Phillips can be reached on 571-272-3940. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

/Hong Cho/

Primary Examiner, Art Unit 2467

RED HAT 1010
RED HAT V. CAS

Notice of References Cited	Application/Control No. 14/512,414	Applicant(s)/Patent Under Reexamination DELANGIS, ERIC M.	
	Examiner HONG CHO	Art Unit 2467	Page 1 of 1

U.S. PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	Classification
*	A	US-2003/0137959	07-2003	Nebiker et al.	370/338
*	B	US-6,496,477	12-2002	Perkins et al.	370/228
*	C	US-7,068,623	06-2006	Barany et al.	370/329
*	D	US-7,257,106	08-2007	Chen et al.	370/338
*	E	US-6,249,669	06-2001	Ogino et al.	455/63.1
*	F	US-2002/0150099	10-2002	Pung et al.	370/390
*	G	US-5,483,522	01-1996	Derby et al.	370/400
	H	US-			
	I	US-			
	J	US-			
	K	US-			
	L	US-			
	M	US-			

FOREIGN PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Country	Name	Classification
	N					
	O					
	P					
	Q					
	R					
	S					
	T					

NON-PATENT DOCUMENTS

*		Include as applicable: Author, Title Date, Publisher, Edition or Volume, Pertinent Pages)
	U	
	V	
	W	
	X	

*A copy of this reference is not being furnished with this Office action. (See MPEP § 707.05(a).)
Dates in MM-YYYY format are publication dates. Classifications may be US or foreign

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	14512414	DELANGIS, ERIC M.
	Examiner	Art Unit
	HONG CHO	2467

CPC- SEARCHED		
Symbol	Date	Examiner

CPC COMBINATION SETS - SEARCHED		
Symbol	Date	Examiner

US CLASSIFICATION SEARCHED			
Class	Subclass	Date	Examiner

SEARCH NOTES		
Search Notes	Date	Examiner
H04L12/5692;H04L45/24;H04W88/16;H04W8/26;H04W28/06;H04W40/02 ;H04W80/00;H04W84/12;H04W84/18;H04W72/04 (EAST text search)	8/4/2015	hc

INTERFERENCE SEARCH			
US Class/ CPC Symbol	US Subclass / CPC Group	Date	Examiner
		8/4/2015	hc

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (**Not for submission under 37 CFR 1.99**)		
	Application Number	14512414
	Filing Date	2014-10-11
	First Named Inventor	Eric Delangis
	Art Unit	2652
	Examiner Name	
	Attorney Docket Number	ED 2000

U.S.PATENTS						Remove
Examiner Initial*	Cite No	Patent Number	Kind Code[1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1					
If you wish to add additional U.S. Patent citation information please click the Add button.						Add

U.S.PATENT APPLICATION PUBLICATIONS						Remove
Examiner Initial*	Cite No	Publication Number	Kind Code[1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1					
If you wish to add additional U.S. Published Application citation information please click the Add button.						Add

FOREIGN PATENT DOCUMENTS							Remove	
Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2] i	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	T[5]
	1							☐
If you wish to add additional Foreign Patent Document citation information please click the Add button							Add	

NON-PATENT LITERATURE DOCUMENTS			Remove
Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc), date, pages(s), volume-issue number(s), publisher, city and/or country where published.	T[5]

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /H.C./

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	14512414 14512414 - GAU: 2467
	Filing Date	2014-10-11
	First Named Inventor	Eric Delangis
	Art Unit	2652
	Examiner Name	
	Attorney Docket Number	ED 2000

	1	Fisk et al., "Dynamic Right-Sizing in TCP", Proceedings of the Los Alamos Computer Science Institute Symposium", pp. 01-5460 (2001)	☐
	2	Hsieh et al., "pTCP: An End-to-End Transport Layer Protocol for Striped Connections", ICNP, pp. 24-33 IEEE Computer Society (2002)	☐

If you wish to add additional non-patent literature document citation information please click the Add button **Add**

EXAMINER SIGNATURE

Examiner Signature	/Hong Cho/	Date Considered	07/28/2015

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /H.C./

RED HAT 1010
RED HAT V. CAS

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	14512414 14512414 - GAU: 2467
	Filing Date	2014-10-11
	First Named Inventor	Eric Delangis
	Art Unit	2652
	Examiner Name	
	Attorney Docket Number	ED 2000

CERTIFICATION STATEMENT

Please see 37 CFR 1.97 and 1.98 to make the appropriate selection(s):

☐ That each item of information contained in the information disclosure statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(1).

OR

☒ That no item of information contained in the information disclosure statement was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the person signing the certification after making reasonable inquiry, no item of information contained in the information disclosure statement was known to any individual designated in 37 CFR 1.56(c) more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(2).

☐ See attached certification statement.

☐ The fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☐ A certification statement is not submitted herewith.

SIGNATURE

A signature of the applicant or representative is required in accordance with CFR 1.33, 10.18. Please see CFR 1.4(d) for the form of the signature.

Signature	/H C McGurk IV/	Date (YYYY-MM-DD)	2014-11-05
Name/Print	H C McGurk	Registration Number	34964

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 1 hour to complete, including gathering, preparing and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /H.C./

RED HAT 1010
RED HAT V. CAS

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these record s.

2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.

3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.

4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).

5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.

6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).

7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.

8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.

9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /H.C./

EAST Search History

EAST Search History (Prior Art)

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
S276	1	14/512414	US-PGPUB; USPAT	OR	OFF	2015/07/30 10:12
S277	12	((Eric) near2 (Delangis)).INV.	US-PGPUB; USPAT; USOCR	OR	OFF	2015/07/30 10:13
S279	3	(connect$3 or establish$3 or join$3) with multi near (link or path) same multicastS3 near router	US-PGPUB; USPAT; USOCR	OR	OFF	2015/07/30 10:18
S281	9	multi near (link or path) same multicast$3 same join$3	US-PGPUB; USPAT; USOCR	OR	OFF	2015/07/30 10:19
S282	1	S277 and connection near id	US-PGPUB; USPAT; USOCR	OR	OFF	2015/07/30 10:20
S287	2	join$3 with multicast$3 near router and multi near (link or path)	US-PGPUB; USPAT; USOCR	OR	OFF	2015/07/30 10:40
S293	7	S292 and @rlad<"20021015"	US-PGPUB; USPAT	OR	OFF	2015/07/30 10:56
S292	166	join$3 with multicast$3 and multi near (path or link)	US-PGPUB; USPAT; USOCR	OR	OFF	2015/07/30 10:56
S295	27	S294 and join$3 with multicastS3 and multi near (path or link)	US-PGPUB; USPAT; USOCR	OR	OFF	2015/07/30 11:12
S294	41830	H04L12/5692;H04L45/24;H04W88/16;H04W8/26;H04W28/06;H04W40/02;H04W80/00;H04W84/12;H04W84/18;H04W72/04.cpc.	US-PGPUB; USPAT	OR	OFF	2015/07/30 11:12
L1	259	(prun$3 or remov$3) with multicast$3 near tree	US-PGPUB; USPAT	OR	OFF	2015/08/11 08:49
L2	32	l1 and @rlad<"20021015"	US-PGPUB; USPAT	OR	OFF	2015/08/11 08:50

8/11/2015 9:27:36 AM
C:\Users\hcho\Documents\EAST\Workspaces\14512414.wsp

RED HAT 1010
RED HAT V. CAS

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

BIB DATA SHEET

CONFIRMATION NO. 9530

SERIAL NUMBER	FILING or 371(c) DATE RULE	CLASS	GROUP ART UNIT	ATTORNEY DOCKET NO.
14/512,414	10/11/2014	370	2467	ED 2000

APPLICANTS

INVENTORS

Eric M. Delangis, Huntsville, AL;

**** CONTINUING DATA ***************************

This application is a CON of 13/531,294 06/22/2012 PAT 8861349 *
which is a CON of 12/581,852 10/19/2009 PAT 8228801
which is a CON of 10/686,375 10/14/2003 PAT 7606156
which claims benefit of 60/418,521 10/15/2002
(*)Data provided by applicant is not consistent with PTO records.

**** FOREIGN APPLICATIONS ***************************

**** IF REQUIRED, FOREIGN FILING LICENSE GRANTED** ** ** SMALL ENTITY **
10/16/2014

Foreign Priority claimed	☐ Yes ☑ No	☐ Met after Allowance	STATE OR COUNTRY	SHEETS DRAWINGS	TOTAL CLAIMS	INDEPENDENT CLAIMS
35 USC 119(a-d) conditions met	☐ Yes ☑ No					
Verified and Acknowledged	/HONG SOL CHO/ Examiner's Signature	Initials	AL	9	20	4

ADDRESS

Harold McGurk
The Law Office of Clay McGurk
P.O. BOX 1488
Orange, CA 92856
UNITED STATES

TITLE

MULTIPATH COMMUNICATION DEVICES AND METHODS

FILING FEE RECEIVED		
940	FEES: Authority has been given in Paper No._____ to charge/credit DEPOSIT ACCOUNT No._____ for following:	☐ All Fees ☐ 1.16 Fees (Filing) ☐ 1.17 Fees (Processing Ext. of time) ☐ 1.18 Fees (Issue) ☐ Other _____ ☐ Credit

BIB (Rev. 05/07).

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (**Not for submission under 37 CFR 1.99**)		
	Application Number	14512414
	Filing Date	2014-10-11
	First Named Inventor	Eric Delangis
	Art Unit	2652
	Examiner Name	
	Attorney Docket Number	ED 2000

U.S.PATENTS						Remove
Examiner Initial*	Cite No	Patent Number	Kind Code[1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1					
If you wish to add additional U.S. Patent citation information please click the Add button.						Add

U.S.PATENT APPLICATION PUBLICATIONS						Remove
Examiner Initial*	Cite No	Publication Number	Kind Code[1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1					
If you wish to add additional U.S. Published Application citation information please click the Add button.						Add

FOREIGN PATENT DOCUMENTS							Remove	
Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2] i	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	T[5]
	1							☐
If you wish to add additional Foreign Patent Document citation information please click the Add button							Add	

NON-PATENT LITERATURE DOCUMENTS			Remove
Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc), date, pages(s), volume-issue number(s), publisher, city and/or country where published.	T[5]

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /H.C./

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	14512414 14512414 - GAU: 2467
	Filing Date	2014-10-11
	First Named Inventor	Eric Delangis
	Art Unit	2652
	Examiner Name	
	Attorney Docket Number	ED 2000

	1	"End-to-End Approach to Host Mobility", by Alex C. Snoeren and Hari Balakrishnan, MIT Laboratory for Computer Science, MobiCom (2000).	☐
	2	"Multi-homed TCP", by Huitema, C. (Internet Draft) IETF, (1995).	☐
	3	"TCP extensions for high performance", by Jacobson V., Braden R., and Borman D., RFC 1323, IETF, (1992).	☐
	4	"A New Multicasting-based Architecture for Internet Host Mobility", by Mysore, J., and Bharghavan, V., Proc. ACM/ IEEE Mobicom '97 (1997).	☐
	5	"Mobile IPv4 Challenge/Response Extensions", by Perkins, C., and Calhoun, P., RFC 3012, IETF, (2000).	☐
	6	"Mobility Support in IPv6", by Perkins, C., and Johnson, D, Proc. ACM/IEEE Mobicom '96 (1996).	☐
	7	"Route Optimization in Mobile IP", by Perkins C., and Johnson, D., Cluster Computing 1, (1998).	☐
	8	"IP Mobility Support", by Perkins, C., RFC 2002, IETF, (1996).	☐
	9	"Transmission Control Protocol", RFC 793, IETF (1981).	☐

If you wish to add additional non-patent literature document citation information please click the Add button **Add**

EXAMINER SIGNATURE

Examiner Signature	/Hong Cho/	Date Considered	07/28/2015

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /H.C./

RED HAT 1010
RED HAT V. CAS

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	14512414 14512414 - GAU: 2467
	Filing Date	2014-10-11
	First Named Inventor	Eric Delangis
	Art Unit	2652
	Examiner Name	
	Attorney Docket Number	ED 2000

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /H.C./

RED HAT 1010
RED HAT V. CAS

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	14512414 14512414 - GAU: 2467
	Filing Date	2014-10-11
	First Named Inventor	Eric Delangis
	Art Unit	2652
	Examiner Name	
	Attorney Docket Number	ED 2000

CERTIFICATION STATEMENT

Please see 37 CFR 1.97 and 1.98 to make the appropriate selection(s):

☐ That each item of information contained in the information disclosure statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(1).

OR

☐ That no item of information contained in the information disclosure statement was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the person signing the certification after making reasonable inquiry, no item of information contained in the information disclosure statement was known to any individual designated in 37 CFR 1.56(c) more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(2).

☐ See attached certification statement.

☐ The fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☐ A certification statement is not submitted herewith.

SIGNATURE

A signature of the applicant or representative is required in accordance with CFR 1.33, 10.18. Please see CFR 1.4(d) for the form of the signature.

Signature	/H C McGurk IV/	Date (YYYY-MM-DD)	2015-07-02
Name/Print	H C McGurk	Registration Number	34964

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 1 hour to complete, including gathering, preparing and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /H.C./

RED HAT 1010
RED HAT V. CAS

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these record s.

2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.

3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.

4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).

5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.

6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).

7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.

8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.

9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /H.C./

Index of Claims	Application/Control No.	Applicant(s)/Patent Under Reexamination
	14512414	DELANGIS, ERIC M.
	Examiner	Art Unit
	HONG CHO	2467

✓	Rejected	-	Cancelled	N	Non-Elected	A	Appeal
=	Allowed	÷	Restricted	I	Interference	O	Objected

☐ Claims renumbered in the same order as presented by applicant ☐ CPA ☐ T.D. ☐ R.1.47

CLAIM		DATE								
Final	Original	08/11/2015								
	1	✓								
	2	✓								
	3	✓								
	4	✓								
	5	✓								
	6	✓								
	7	✓								
	8	✓								
	9	✓								
	10	✓								
	11	✓								
	12	✓								
	13	✓								
	14	✓								
	15	✓								
	16	✓								
	17	✓								
	18	✓								
	19	✓								
	20	✓								

To: CLAYMCGURK@GMAIL.COM,,
From: PAIR_eOfficeAction@uspto.gov
Cc: PAIR_eOfficeAction@uspto.gov
Subject: Private PAIR Correspondence Notification for Customer Number 64828

Aug 18, 2015 05:21:54 AM

Dear PAIR Customer:

Harold McGurk
The Law Office of Clay McGurk
P.O. BOX 1488
Orange, CA 92856
UNITED STATES

The following USPTO patent application(s) associated with your Customer Number, 64828 , have new outgoing correspondence. This correspondence is now available for viewing in Private PAIR.

The official date of notification of the outgoing correspondence will be indicated on the form PTOL-90 accompanying the correspondence.

Disclaimer:
The list of documents shown below is provided as a courtesy and is not part of the official file wrapper. The content of the images shown in PAIR is the official record.

Application	Document	Mailroom Date	Attorney Docket No.
14512414	CTNF	08/18/2015	ED 2000
	892	08/18/2015	ED 2000
	1449	08/18/2015	ED 2000
	1449	08/18/2015	ED 2000

To view your correspondence online or update your email addresses, please visit us anytime at https://sportal.uspto.gov/secure/myportal/privatepair.

If you have any questions, please email the Electronic Business Center (EBC) at EBC@uspto.gov with 'e-Office Action' on the subject line or call 1-866-217-9197 during the following hours:

Monday - Friday 6:00 a.m. to 12:00 a.m.

Thank you for prompt attention to this notice,

UNITED STATES PATENT AND TRADEMARK OFFICE

PATENT APPLICATION INFORMATION RETRIEVAL SYSTEM

RED HAT 1010
RED HAT V. CAS

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION NO.:	**14/512,414**	EXAMINER:	HONG CHO
INVENTOR:	Eric M. Delangis	GROUP ART UNIT:	2467
FILED:	OCTOBER 11, 2014	CONFIRMATION NO.:	9530
TITLE:	Multipath Communication Devices and Methods		
ATTORNEY NO:	ED 2000		

Commissioner For Patents
PO Box 1450
Alexandria, VA 22313-1450

AMENDMENT

In response to the Office Action issued on August 18, 2015 ("Office Action"), please reconsider the above-identified application in light of the following amendments and accompanying remarks. An Information Disclosure Statement is also being concurrently submitted with this Amendment. This Amendment, IDS and the requisite fees are being submitted on November 18, 2015.

CERTIFICATE OF ELECTRONIC (EFS-WEB) TRANSMISSION
I hereby certify that this correspondence is being transmitted via the U.S. Patent Office's electronic filing system in accordance with 37 C.F.R. § 1.8(a)(1) from the **Pacific Time Zone** of the United States on the local date shown below.

Dated: November 18, 2015 By: Clay McGurk
Clay McGurk

RED HAT 1010
RED HAT V. CAS

CLAIMS

This listing of claims will replace all prior versions and listings of the claims in the application:

1. (Currently Amended) A method for establishing and sending data during a multipath connection ~~across multiple paths between a host and an initiating device, the host and a plurality of devices capable of being connected to one or more networks,~~ comprising:

a data-source device receiving ~~at the host~~ multipath information about each of a plurality of network-edge devices that ~~are~~ is capable of joining a multipath connection with a data-~~requesting the initiating~~ device, each of the network-edge devices being connected to one or more networks, and each of the network-edge devices being connected to the data-requesting device over a different connection; and

the data-source device ~~host~~ performing at least one of (i) sending different groups of data packets to ~~one or more of the~~ different network-edge devices that are participating in the multipath connection, (ii) sending ~~separate~~ different streams of data to one or more of the network-edge devices that are participating in the multipath connection, and (iii) sending combinations of ~~separate~~ different streams and different groups of data packets to one or more of the network-edge devices that are participating in the multipath connection.

2. (Currently Amended) The method of claim 1, wherein the multipath information includes a connection ID that is generated by one of the ~~host~~ data-source device, one of the plurality of network-edge devices, the ~~initiating~~ data-requesting device and a network device.

3. (Currently Amended) The method of claim ~~2~~1, further comprising:

the data-source device receiving ~~at the host~~ second multipath information about ~~one or more additional devices~~ a second network-edge device that is capable of joining the multipath connection;[[,]] and

the data-source device joining the ~~one or more additional devices~~ second network-edge device ~~being associated with the connection ID~~ to the multipath connection.

RED HAT 1010
RED HAT V. CAS

4. (Currently Amended) The method of claim ~~2~~1, further comprising:

the ~~host~~ data-source device receiving ~~at the host~~ multipath information about which network-edge devices are to be removed from the multipath connection[[,]]; and

the data-source device removing the network-edge devices ~~to be removed~~ from being associated with the ~~connection ID~~ multipath connection.

5. (Currently Amended) The method of claim 1, further comprising at least one of the ~~host~~ data-source device, the network-edge devices and the ~~initiating~~ data-requesting device determining which of the network-edge devices are to be removed from the multipath connection.

6. (Currently Amended) The method of claim 1, wherein the ~~host~~ data-source device is one of a multipath link aggregator, a network relay, a multipath link server, a server, a network switch, a network router and a device capable of providing multipath services.

7. (Currently Amended) A method for establishing and sending data ~~across multiple paths in different multipath connections from a host~~ during a multipath connection, comprising:

the data-source device ~~host~~ receiving multipath connection requests from a plurality of data-requesting devices, each of the multipath connection requests involving a particular ~~different~~ multipath connection to a data-requesting device;

the data-source device ~~host~~ associating a connection ID ~~for~~ to each multipath connection request;

the data-source device using ~~host receiving~~ multipath information ~~of~~ about each of a plurality of network-edge devices that is capable of joining in a particular multipath connection to a particular data-requesting device, the particular multipath connection having a particular connection ID;

the data-source device notifying each of the network-edge devices that is participating in the particular multipath connection associated with the particular connection ID; and

the data-source device ~~host~~ sending data associated with the particular connection ID to ~~one or more~~ ~~of~~ the network-edge devices participating in the particular multipath connection associated with the particular connection ID.

RED HAT 1010
RED HAT V. CAS

8. (Currently Amended) The method of claim 7, further comprising: ~~receiving at the host information about one or more devices capable of joining the particular multipath connection, the one or more devices being associated with the particular connection ID~~ the data-source device sending different groups of data packets to different network-edge devices that are participating in the particular multipath connection associated with the particular connection ID.

9. (Currently Amended) The method of claim 7, further comprising:

the data-source device receiving ~~at the host~~ identification of the network-edge devices to be removed from the particular multipath connection[[,]]; and

the data-source device removing the identified network-edge devices ~~to be removed being associated with the particular connection ID~~ from the particular multipath connection.

10. (Currently Amended) The method of claim 7, further comprising ~~at least one of the host and the devices determining which of the devices are to be removed from the particular multipath connection~~ the data-source device notifying the data-requesting device when all data has been sent over the particular multilink connection.

11. (Currently Amended) The method of claim 7, wherein the data-source device ~~host~~ is one of a multipath link aggregator, a network relay, a multipath link server, a server, a network switch, a network router and a device capable of providing multipath services.

12. (Currently Amended) The method of claim 7, wherein the connection ID is generated by one of the data-source ~~host~~ device, one of the plurality of network-edge devices, the data-requesting device, and a network device.

RED HAT 1010
RED HAT V. CAS

13. (Currently Amended) A method for establishing and sending data during a multipath connection, ~~across multiple paths between a device and a host, the device capable of being connected to at least one of a network and one or more network devices, the remote devices capable of being connected to at least one network, the host capable of being connected to one or more networks,~~ comprising:

a data-requesting device sending a multipath request to a network-edge device where the multipath request indicates that the network-edge device will join a multipath connection associated with the data-requesting device, the data-requesting device being connected over a first connection to a network and over a second connection to the network-edge device, the multipath connection comprising the first connection and the second connection which are capable of receiving data at the same time; ~~one or more network devices capable of joining a multipath connection with the device; and~~

the data-requesting device receiving first data ~~from the host at the device~~ over the first connection from the network; and ~~via one or more of the identified network devices during the multipath connection~~

after the network-edge device joins the multipath connection, the data-requesting device receiving second data over the second connection from the network-edge device.

14. (Currently Amended) The method of claim 13, ~~further comprising receiving at the device information about the one or more network devices, the information including a confirmation of a capability of joining in the multipath connection~~ wherein the first connection and the second connection are each a wireless connection, a wired connection or a combination of wired and wireless connections.

15. (Currently Amended) The method of claim 13, wherein the ~~network~~ data-requesting device is one of a cell phone, a tablet, a switch, a network device, a router, a computer, a Wi-Fi device, a Wi-Max device, a communications device, a mobile communications device, a near-field communications device, and a Bluetooth device.

RED HAT 1010
RED HAT V. CAS

16. (Currently Amended) The method of claim 13, wherein the network-edge ~~devices comprise~~ device comprises one or more routers, modems, cell phones, tablets, computers, cellular towers, wireless towers, base stations, wireless base stations, repeaters, wireless repeaters, Wi-Fi devices, networking devices, hot spot devices, Wi-MAX devices, LTE devices, GSM devices, consumer electronic devices, an Internet-ready device, a network switch, a network router, a multipath server and a device capable of providing multipath services.

17. (Currently Amended) The method of claim 13, wherein the ~~selected one or more network devices~~ data-requesting device is capable of connecting to another network device which is capable of joining the multipath connection, and further comprising the data-requesting device receiving ~~relaying~~ data from the another network device ~~to the device through the selected one or more network devices~~ after the another network device joins the multipath connection.

18. (Currently Amended) A ~~communications~~ data-requesting device comprising:

at least one interface capable of connecting concurrently to ~~at least one of~~ a plurality of network-edge devices during a multipath connection, the multipath connection comprising the connections between the data-requesting device and each of the network-edge devices; and

at least one ~~a~~ processor coupled to the at least one interface that

~~identifies one or more network devices capable of joining a multipath connection with the communications device;~~

sends multipath information to each of the network-edge devices that is capable of joining the multipath connection; and

receives different data at the data-requesting device ~~via one or more of the identified~~ from different ones of the network-edge devices ~~during~~ that are able to join the multipath connection.

19. (Currently Amended) The ~~communications~~ data-requesting device of claim 18, wherein the data-requesting device is one of a cell phone, a tablet, a switch, a network device, a router, a computer, a Wi-Fi device, a Wi-Max device, a mobile communications device, a near-field communications device, and a Bluetooth device.

RED HAT 1010
RED HAT V. CAS

20. (Currently Amended) The ~~communications~~ data-requesting device of claim 18, wherein each of the network-edge devices comprises one or more network devices, communications devices, cellular devices, tablet devices, router devices, computer devices, satellite devices, Wi-MAX devices, wireless devices, mobile communications devices, Bluetooth devices, and Wi-Fi devices, a network switch, a network router, a multipath server and a device capable of providing multipath services.

21. (New) A method for creating and using a multipath connection, comprising the steps of:

a data-requesting device sending first multipath control information to or through a first network-edge device, the first multipath control information indicates that the first network-edge device will join the multipath connection, the data-requesting device being connected over a first connection to the first network-edge device;

the data-requesting device sending a second multipath control information to or through a second network-edge device, the second multipath control information indicates that the second network-edge device will join the multipath connection, the data-requesting device being connected over a second connection to the second network-edge device, the data-requesting device capable of receiving different data over the first connection and the second connection at the same time; and

after the first network-edge device joins the multipath connection, the data-requesting device receiving first data over the first connection from the first network-edge device; and

after the second network-edge device joins the multipath connection, receiving second data at the data-requesting device over the second path from the second network-edge device.

22. (New) The method as in claim 21, wherein the first network-edge device is connected to a network, and wherein the second network-edge device is connected to the network or a different network.

23. (New) The method as in claim 21, wherein the first connection and the second connection are wireless connections, a wired connection or a combination of wired and wireless connections.

RED HAT 1010
RED HAT V. CAS

REMARKS

Status of the Claims

Claims 1-20 were pending in this application before this Amendment. Claims 1-20 have been amended for clarifying terminology of what device is performing what function or functions. Claims 21-23 are new claims. No new matter has been added in this Amendment. Thus, claims 1-23 are now pending pursuant to this Amendment.

The claims have been amended to recite "data-source device", "data-requesting device" and "network-edge device". It is hoped that this terminology will better aid in understanding what is being claimed. As shown in figures 7-9 and the accompanying description of the present application, the "data-requesting device" is a device (e.g., computer, tablet, cell phone, router) requesting data over the network from the "data-source device" (e.g., server) using two or more connections to different "network-edge device" during a multipath connection. The "data-source device" provides data to a data-requesting device over the network through different "network-edge devices" during a multipath connection.

35 U.S.C. § 102 Rejections

Starting on page 2 of the Office Action, claims 1-3, 6-8 and 11-20 were rejected under 35 U.S.C. §102(e) as being unpatentable over "Perkins" (U.S. Patent 6,496,477). It is easy to see the differences between Perkins and the amended independent claims 1, 7, 13 and 18 by looking at Fig. 1 in Perkins. Each of the "data-requesting devices" of Perkins, computer 105, cell phones 181 and 189 only have one connection to the network 100. For computer 105, the connection to the network is through modem 168.i; for cell phone 181, the connection to the network 100 is through cell tower 183; and, for cell phone 187, the connection is through cell tower 187. All the amended independent claims 1, 7, 13 and 18 require at least two different "network-edge devices" during a multipath connection which are connected to the data-requesting device. Perkins completely lacks the required two different network-edge devices.

Figures 11 and 12 of Perkins show a single connection between a network-edge device and a data-requesting device. The whole purpose of Perkins is multicasting a stream of data into multiple streams of data, which is clearly illustrated in Fig. 12. SRC 1111 provides one stream of data, which is multicast by nodes 1125 and 1135 into multiple streams to different network nodes, such as network node 1131. Although node 1131 has two different streams multicast to

it, node 1131 is a part of the network 100. That corresponds to Fig. 1 of Perkins where cell towers 183 and 187 receive data from PSTN 185 and also gateway 191 and 193, respectively. The data-requesting device still only has one path to a network-edge device (e.g., cell tower or modem) and not two as required for a multipath connection for amended independent claims 1, 7, 13 and 18. Moreover, the data streamed by nodes 1125 and 1135 is the same data, while the claims have been amended to show that different data is sent over different connections from the network-edge devices to the data-requesting device.

For these reasons, claims 1-3, 6-8 and 11-20 are patentably distinct and not anticipated by Perkins. It is respectfully requested that the rejection of claims 1-3, 6-8 and 11-20 under 35 U.S.C. §102 based on Perkins be reconsidered and withdrawn.

35 U.S.C. § 103 Rejections

Starting on page 4 of the Office Action, claims 4, 5, 9 and 10 were rejected under 35 U.S.C. §102(e) as being unpatentable over Perkins in view of "Pung" (U.S. Publication 2002-0150099). Pung suffers from the same deficiencies as Perkins as shown in Fig. 1 of Pung, in that there is one only connection between the network and the data-requesting device. Claims 4, 5, 9 and 10 are allowable for the same reasons as provided above for their corresponding amended independent claim. For this reason and those reasons distinguishing Perkins, it is respectfully requested that the rejection of claims 4, 5, 9 and 10 under 35 U.S.C. § 103 based on Perkins and Pung be reconsidered and withdrawn.

Conclusion

It is noted that claimed subject matter may be patentably distinguished from the cited references for additional reasons. It is also noted that any failure to comment directly upon any of the positions asserted in the Office Action does not indicate agreement or acquiescence with those asserted positions or allegations. Instead, it is believed that the foregoing sufficiently addresses the issues raised in the Office Action, and that the present claims are in condition for allowance.

All rejections in the Office Action have been properly traversed. It is respectfully requested that all outstanding rejections be reconsidered and withdrawn. It is believed that a full

RED HAT 1010
RED HAT V. CAS

and complete reply has been made in response to the Office Action, and, that as such, the present patent application is in a condition for allowance.

The Examiner is always welcome to telephone the attorney of record at the telephone number provided below if such telephone discussion would help expedite the prosecution of the patent application. Prompt and favorable consideration of this Amendment is respectfully requested.

Respectfully Submitted,

Date: November 18, 2015

H C McGurk IV, Reg. No. 34,964
Attorney for Applicant

The Law Office of Clay McGurk
PO Box 1488
Orange, CA 92706
Phone: **949.395.3942**
Email: claymcgurk@gmail.com

RED HAT 1010
RED HAT V. CAS

PTO/SB/08a (01-10)
Approved for use through 07/31/2012. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Not for submission under 37 CFR 1.99)

Application Number	14512414
Filing Date	2014-10-11
First Named Inventor	Eric Delangis
Art Unit	2467
Examiner Name	Hong Cho
Attorney Docket Number	ED 2000

U.S.PATENTS Remove

Examiner Initial*	Cite No	Patent Number	Kind Code[1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1	7224964		2007-05-29	Souissi et al.	
	2	7917129		2011-03-29	Souisse et al.	
	3	8442504		2013-05-14	Souissi et al.	
	4	8521142		2013-08-27	Souissi et al.	

If you wish to add additional U.S. Patent citation information please click the Add button. Add

U.S.PATENT APPLICATION PUBLICATIONS Remove

Examiner Initial*	Cite No	Publication Number	Kind Code[1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1					

If you wish to add additional U.S. Published Application citation information please click the Add button. Add

FOREIGN PATENT DOCUMENTS Remove

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2] i	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	T[5]

EFS Web 2.1.17

RED HAT 1010
RED HAT V. CAS

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	14512414
	Filing Date	2014-10-11
	First Named Inventor	Eric Delangis
	Art Unit	2467
	Examiner Name	Hong Cho
	Attorney Docket Number	ED 2000

	1							☐

If you wish to add additional Foreign Patent Document citation information please click the Add button Add

NON-PATENT LITERATURE DOCUMENTS Remove

Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc), date, pages(s), volume-issue number(s), publisher, city and/or country where published.	T[5]
	1	"Multiple Paths Protocol for a Cluster Type Network", Hiroshi Mineno et al., 7 pages (1999)	☐
	2	"The PPP Multilink Protocol (MP)", RFC 1990, 24 pages (1996)	☐

If you wish to add additional non-patent literature document citation information please click the Add button Add

EXAMINER SIGNATURE

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	14512414
	Filing Date	2014-10-11
	First Named Inventor	Eric Delangis
	Art Unit	2467
	Examiner Name	Hong Cho
	Attorney Docket Number	ED 2000

CERTIFICATION STATEMENT

Please see 37 CFR 1.97 and 1.98 to make the appropriate selection(s):

☐ That each item of information contained in the information disclosure statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(1).

OR

☐ That no item of information contained in the information disclosure statement was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the person signing the certification after making reasonable inquiry, no item of information contained in the information disclosure statement was known to any individual designated in 37 CFR 1.56(c) more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(2).

☐ See attached certification statement.

☒ The fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☐ A certification statement is not submitted herewith.

SIGNATURE

A signature of the applicant or representative is required in accordance with CFR 1.33, 10.18. Please see CFR 1.4(d) for the form of the signature.

Signature	/H C McGurk IV/	Date (YYYY-MM-DD)	2015-11-18
Name/Print	H C McGurk	Registration Number	34964

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 1 hour to complete, including gathering, preparing and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these record s.

2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.

3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.

4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).

5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.

6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).

7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.

8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.

9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION NO.:	**14/512,414**	EXAMINER:	HONG CHO
INVENTOR:	Eric M. Delangis	GROUP ART UNIT:	2467
FILED:	OCTOBER 11, 2014	CONFIRMATION NO.:	9530
TITLE:	Multipath Communication Devices and Methods		
ATTORNEY NO:	ED 2000		

Commissioner For Patents
PO Box 1450
Alexandria, VA 22313-1450

INFORMATION DISCLOSURE STATEMENT TRANSMITTAL LETTER

This submission does not represent that a search has been made or that no better art exists and does not constitute an admission that each or all of the listed documents are material or constitute prior art. If the Examiner applies any of the documents as prior art against any of the claims in the present application, all rights are reserved to take appropriate action to establish the patentability of the disclosed invention over the present application.

Best Regards,

Date: November 18, 2015

H C McGurk IV, Reg. No. 34,964
Attorney for Applicant

The Law Office of Clay McGurk
PO Box 1488
Orange, CA 92856
Phone: **949.395.3942**

CERTIFICATE OF ELECTRONIC (EFS-WEB) TRANSMISSION

I hereby certify that this correspondence is being transmitted via the U.S. Patent Office's electronic filing system in accordance with 37 C.F.R. § 1.8(a)(1) from the **Pacific Time Zone** of the United States on the local date shown below.

Dated: November 18, 2015

By: Clay McGurk

RED HAT 1010
RED HAT V. CAS

Network Working Group
Request for Comments: 1990
Obsoletes: 1717
Category: Standards Track

K. Sklower
University of California, Berkeley
B. Lloyd
G. McGregor
Lloyd Internetworking
D. Carr
Newbridge Networks Corporation
T. Coradetti
Sidewalk Software

August 1996

The PPP Multilink Protocol (MP)

Status of this Memo

This document specifies an Internet standards track protocol for the Internet community, and requests discussion and suggestions for improvements. Please refer to the current edition of the "Internet Official Protocol Standards" (STD 1) for the standardization state and status of this protocol. Distribution of this memo is unlimited.

Abstract

This document proposes a method for splitting, recombining and sequencing datagrams across multiple logical data links. This work was originally motivated by the desire to exploit multiple bearer channels in ISDN, but is equally applicable to any situation in which multiple PPP links connect two systems, including async links. This is accomplished by means of new PPP [2] options and protocols.

The differences between the current PPP Multilink specification (RFC 1717) and this memo are explained in Section 11. Any system implementing the additional restrictions required by this memo will be backwards compatible with conforming RFC 1717 implementations.

Acknowledgements

The authors specifically wish to thank Fred Baker of ACC, Craig Fox of Network Systems, Gerry Meyer of Spider Systems, Dan Brennan of Penril Datability Networks, Vernon Schryver of SGI (for the comprehensive discussion of padding), and the members of the IP over Large Public Data Networks and PPP Extensions working groups, for much useful discussion on the subject.

RED HAT 1010
RED HAT V. CAS

Table of Contents

1. Introduction 2
1.1. Motivation 2
1.2. Functional Description 3
1.3. Conventions 4
2. General Overview 4
3. Packet Formats 7
3.1. Padding Considerations 10
4. Trading Buffer Space Against Fragment Loss 10
4.1. Detecting Fragment Loss 11
4.2. Buffer Space Requirements 12
5. PPP Link Control Protocol Extensions 13
5.1. Configuration Option Types 13
5.1.1. Multilink MRRU LCP option 14
5.1.2. Short Sequence Number Header Format Option 15
5.1.3. Endpoint Discriminator Option 15
6. Initiating use of Multilink Headers 19
7. Closing Member links 20
8. Interaction with Other Protocols 20
9. Security Considerations 21
10. References 21
11. Differences from RFC 1717 22
11.1. Negotiating Multilink, per se 22
11.2. Initial Sequence Number defined 22
11.3. Default Value of the MRRU 22
11.4. Config-Nak of EID prohibited 22
11.5. Uniformity of Sequence Space 22
11.6. Commencing and Abating use of Multilink Headers 23
11.7. Manual Configuration and Bundle Assignment 23
12. Authors' Addresses 24

1. Introduction

1.1. Motivation

Basic Rate and Primary Rate ISDN both offer the possibility of opening multiple simultaneous channels between systems, giving users additional bandwidth on demand (for additional cost). Previous proposals for the transmission of internet protocols over ISDN have stated as a goal the ability to make use of this capability, (e.g., Leifer et al., [1]).

There are proposals being advanced for providing synchronization between multiple streams at the bit level (the BONDING proposals); such features are not as yet widely deployed, and may require additional hardware for end system. Thus, it may be useful to have a purely software solution, or at least an interim measure.

RED HAT 1010
RED HAT V. CAS

There are other instances where bandwidth on demand can be exploited, such as using a dialup async line at 28,800 baud to back up a leased synchronous line, or opening additional X.25 SVCs where the window size is limited to two by international agreement.

The simplest possible algorithms of alternating packets between channels on a space available basis (which might be called the Bank Teller's algorithm) may have undesirable side effects due to reordering of packets.

By means of a four-byte sequencing header, and simple synchronization rules, one can split packets among parallel virtual circuits between systems in such a way that packets do not become reordered, or at least the likelihood of this is greatly reduced.

1.2. Functional Description

The method discussed here is similar to the multilink protocol described in ISO 7776 [4], but offers the additional ability to split and recombine packets, thereby reducing latency, and potentially increase the effective maximum receive unit (MRU). Furthermore, there is no requirement here for acknowledged-mode operation on the link layer, although that is optionally permitted.

Multilink is based on an LCP option negotiation that permits a system to indicate to its peer that it is capable of combining multiple physical links into a "bundle". Only under exceptional conditions would a given pair of systems require the operation of more than one bundle connecting them.

Multilink is negotiated during the initial LCP option negotiation. A system indicates to its peer that it is willing to do multilink by sending the multilink option as part of the initial LCP option negotiation. This negotiation indicates three things:

1. The system offering the option is capable of combining multiple physical links into one logical link;

2. The system is capable of receiving upper layer protocol data units (PDU) fragmented using the multilink header (described later) and reassembling the fragments back into the original PDU for processing;

3. The system is capable of receiving PDUs of size N octets where N is specified as part of the option even if N is larger than the maximum receive unit (MRU) for a single physical link.

RED HAT 1010
RED HAT V. CAS

Once multilink has been successfully negotiated, the sending system is free to send PDUs encapsulated and/or fragmented with the multilink header.

1.3. Conventions

The following language conventions are used in the items of specification in this document:

- MUST, SHALL or MANDATORY -- the item is an absolute requirement of the specification.
- SHOULD or RECOMMENDED -- the item should generally be followed for all but exceptional circumstances.
- MAY or OPTIONAL -- the item is truly optional and may be followed or ignored according to the needs of the implementor.

2. General Overview

In order to establish communications over a point-to-point link, each end of the PPP link must first send LCP packets to configure the data link during Link Establishment phase. After the link has been established, PPP provides for an Authentication phase in which the authentication protocols can be used to determine identifiers associated with each system connected by the link.

The goal of multilink operation is to coordinate multiple independent links between a fixed pair of systems, providing a virtual link with greater bandwidth than any of the constituent members. The aggregate link, or bundle, is named by the pair of identifiers for two systems connected by the multiple links. A system identifier may include information provided by PPP Authentication [3] and information provided by LCP negotiation. The bundled links can be different physical links, as in multiple async lines, but may also be instances of multiplexed links, such as ISDN, X.25 or Frame Relay. The links may also be of different kinds, such as pairing dialup async links with leased synchronous links.

We suggest that multilink operation can be modeled as a virtual PPP link-layer entity wherein packets received over different physical link-layer entities are identified as belonging to a separate PPP network protocol (the Multilink Protocol, or MP) and recombined and sequenced according to information present in a multilink fragmentation header. All packets received over links identified as belonging to the multilink arrangement are presented to the same network-layer protocol processing machine, whether they have multilink headers or not.

RED HAT 1010
RED HAT V. CAS

The packets to be transmitted using the multilink procedure are encapsulated according to the rules for PPP where the following options would have been manually configured:

- No async control character Map
- No Magic Number
- No Link Quality Monitoring
- Address and Control Field Compression
- Protocol Field Compression
- No Compound Frames
- No Self-Describing-Padding

According to the rules specified in RFC1661, this means that an implementation MUST accept reassembled packets with and without leading zeroes present in the Protocol Field of the reassembled packet. Although it is explicitly forbidden below to include the Address and Control fields (usually, the two bytes FF 03) in the material to be fragmented, it is a good defensive programming practice to accept the packet anyway, ignoring the two bytes if present, as that is what RFC1661 specifies.

As a courtesy to implementations that perform better when certain alignment obtains, it is suggested that a determination be made when a bundle is created on whether to transmit leading zeroes by examining whether PFC has been negotiated on the first link admitted into a bundle. This determination should be kept in force so long as a bundle persists.

Of course, individual links are permitted to have different settings for these options. As described below, member links SHOULD negotiate Self-Describing-Padding, even though pre-fragmented packets MUST NOT be padded. Since the Protocol Field Compression mode on the member link allows a sending system to include a leading byte of zero or not at its discretion, this is an alternative mechanism for generating even-length packets.

LCP negotiations are not permitted on the bundle itself. An implementation MUST NOT transmit LCP Configure-Request, -Reject, -Ack, -Nak, Terminate-Request or -Ack packets via the multilink procedure, and an implementation receiving them MUST silently discard them. (By "silently discard" we mean to not generate any PPP packets in response; an implementation is free to generate a log entry registering the reception of the unexpected packet). By contrast, other LCP packets having control functions not associated with changing the defaults for the bundle itself are permitted. An implementation MAY transmit LCP Code-Reject, Protocol-Reject, Echo-Request, Echo-Reply and Discard-Request Packets.

RED HAT 1010
RED HAT V. CAS

The effective MRU for the logical-link entity is negotiated via an LCP option. It is irrelevant whether Network Control Protocol packets are encapsulated in multilink headers or not, or even over which link they are sent, once that link identifies itself as belonging to a multilink arrangement.

Note that network protocols that are not sent using multilink headers cannot be sequenced. (And consequently will be delivered in any convenient way).

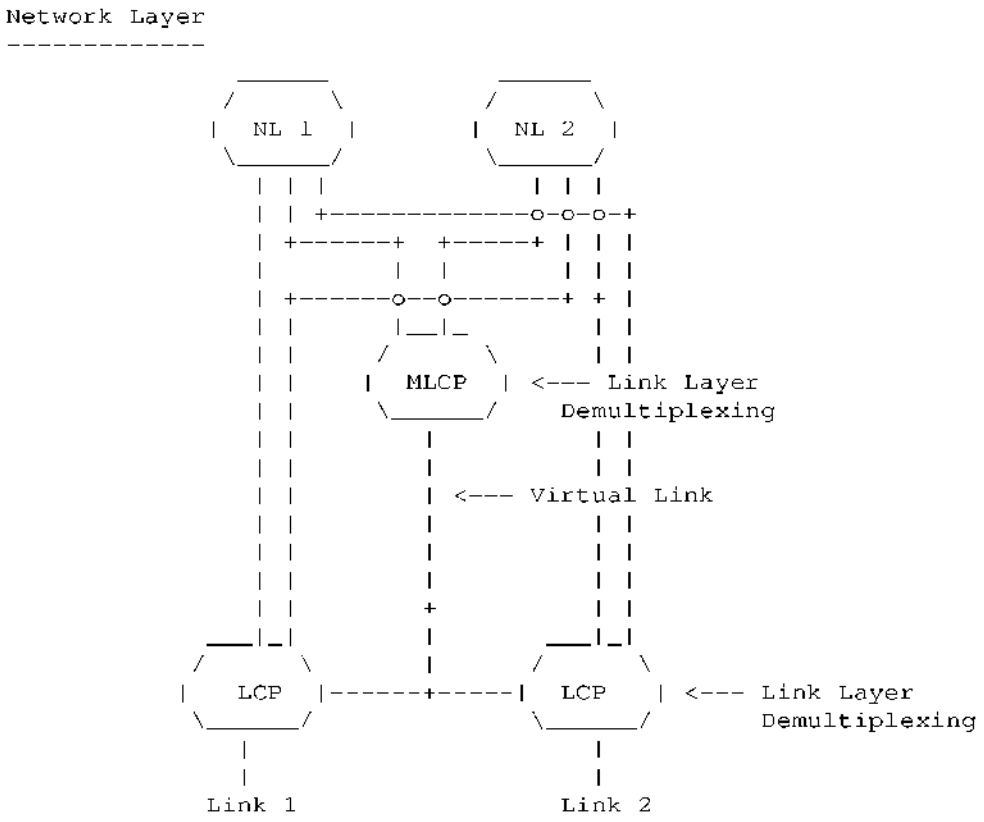
For example, consider the case in Figure 1. Link 1 has negotiated network layers NL 1, NL 2, and MP between two systems. The two systems then negotiate MP over Link 2.

Frames received on link 1 are demultiplexed at the data link layer according the PPP network protocol identifier and can be sent to NL 1, NL 2, or MP. Link 2 will accept frames with all network protocol identifiers that Link 1 does.

Frames received by MP are further demultiplexed at the network layer according to the PPP network protocol identifier and sent to NL 1 or NL 2. Any frames received by MP for any other network layer protocols are rejected using the normal protocol reject mechanism.

RED HAT 1010
RED HAT V. CAS

Figure 1. Multilink Overview.



3. Packet Formats

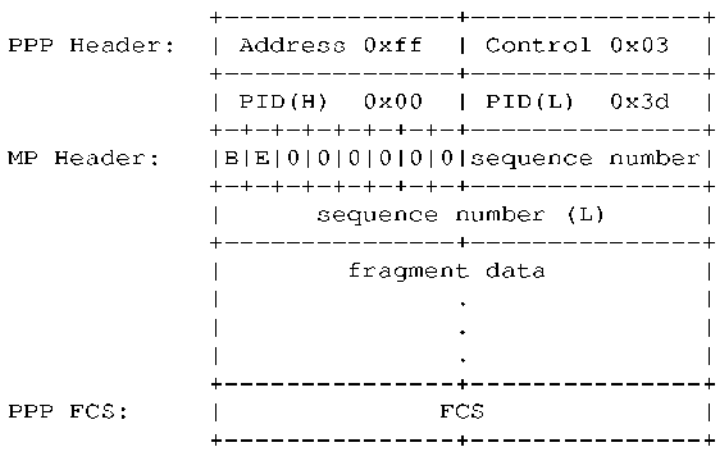
In this section we describe the layout of individual fragments, which are the "packets" in the Multilink Protocol. Network Protocol packets are first encapsulated (but not framed) according to normal PPP procedures, and large packets are broken up into multiple segments sized appropriately for the multiple physical links. Although it would otherwise be permitted by the PPP spec, implementations MUST NOT include the Address and Control Field in the logical entity to be fragmented. A new PPP header consisting of the Multilink Protocol Identifier, and the Multilink header is inserted before each section. (Thus the first fragment of a multilink packet in PPP will have two headers, one for the fragment, followed by the header for the packet itself).

RED HAT 1010
RED HAT V. CAS

Systems implementing the multilink procedure are not required to fragment small packets. There is also no requirement that the segments be of equal sizes, or that packets must be broken up at all. A possible strategy for contending with member links of differing transmission rates would be to divide the packets into segments proportion to the transmission rates. Another strategy might be to divide them into many equal fragments and distribute multiple fragments per link, the numbers being proportional to the relative speeds of the links.

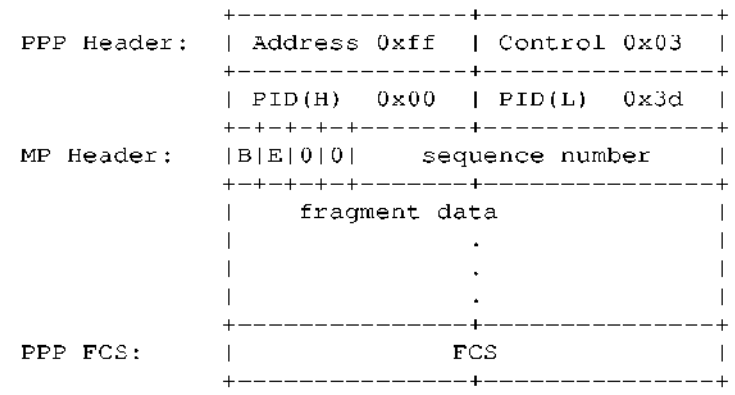
PPP multilink fragments are encapsulated using the protocol identifier 0x00-0x3d. Following the protocol identifier is a four byte header containing a sequence number, and two one bit fields indicating that the fragment begins a packet or terminates a packet. After negotiation of an additional PPP LCP option, the four byte header may be optionally replaced by a two byte header with only a 12 bit sequence space. Address & Control and Protocol ID compression are assumed to be in effect. Individual fragments will, therefore, have the following format:

Figure 2: Long Sequence Number Fragment Format.



RED HAT 1010
RED HAT V. CAS

Figure 3: Short Sequence Number Fragment Format.



The (B)eginning fragment bit is a one bit field set to 1 on the first fragment derived from a PPP packet and set to 0 for all other fragments from the same PPP packet.

The (E)nding fragment bit is a one bit field set to 1 on the last fragment and set to 0 for all other fragments. A fragment may have both the (B)eginning and (E)nding fragment bits set to 1.

The sequence field is a 24 bit or 12 bit number that is incremented for every fragment transmitted. By default, the sequence field is 24 bits long, but can be negotiated to be only 12 bits with an LCP configuration option described below.

Between the (E)nding fragment bit and the sequence number is a reserved field, whose use is not currently defined, which MUST be set to zero. It is 2 bits long when the use of short sequence numbers has been negotiated, 6 bits otherwise.

In this multilink protocol, a single reassembly structure is associated with the bundle. The multilink headers are interpreted in the context of this structure.

The FCS field shown in the diagram is inherited from the normal framing mechanism from the member link on which the packet is transmitted. There is no separate FCS applied to the reconstituted packet as a whole if transmitted in more than one fragment.

RED HAT 1010
RED HAT V. CAS

3.1. Padding Considerations

Systems that support the multilink protocol SHOULD implement Self-Describing-Padding. A system that implements self-describing-padding by definition will either include the padding option in its initial LCP Configure-Requests, or (to avoid the delay of a Configure-Reject) include the padding option after receiving a NAK containing the option.

A system that must pad its own transmissions but does not use Self-Describing-Padding when not using multilink, MAY continue to not use Self-Describing-Padding if it ensures by careful choice of fragment lengths that only (E)nding fragments of packets are padded. A system MUST NOT add padding to any packet that cannot be recognized as padded by the peer. Non-terminal fragments MUST NOT be padded with trailing material by any other method than Self-Describing-Padding.

A system MUST ensure that Self-Describing-Padding as described in RFC 1570 [11] is negotiated on the individual link before transmitting any multilink data packets if it might pad non-terminal fragments or if it would use network or compression protocols that are vulnerable to padding, as described in RFC 1570. If necessary, the system that adds padding MUST use LCP Configure-NAK's to elicit a Configure-Request for Self-Describing-Padding from the peer.

Note that LCP Configure-Requests can be sent at any time on any link, and that the peer will always respond with a Configure-Request of its own. A system that pads its transmissions but uses no protocols other than multilink that are vulnerable to padding MAY delay ensuring that the peer has Configure-Requested Self-Describing-Padding until it seems desireable to negotiate the use of Multilink itself. This permits the interoperability of a system that pads with older peers that support neither Multilink nor Self-Describing-Padding.

4. Trading Buffer Space Against Fragment Loss

In a multilink procedure one channel may be delayed with respect to the other channels in the bundle. This can lead to fragments being received out of order, thus increasing the difficulty in detecting the loss of a fragment. The task of estimating the amount of space required for buffering on the receiver becomes more complex because of this. In this section we discuss a technique for declaring that a fragment is lost, with the intent of minimizing the buffer space required, yet minimizing the number of avoidable packet losses.

RED HAT 1010
RED HAT V. CAS

4.1. Detecting Fragment Loss

On each member link in a bundle, the sender MUST transmit fragments with strictly increasing sequence numbers (modulo the size of the sequence space). This requirement supports a strategy for the receiver to detect lost fragments based on comparing sequence numbers. The sequence number is not reset upon each new PPP packet, and a sequence number is consumed even for those fragments which contain an entire PPP packet, i.e., one in which both the (B)eginning and (E)nding bits are set.

An implementation MUST set the sequence number of the first fragment transmited on a newly-constructed bundle to zero. (Joining a secondary link to an exisiting bundle is invisible to the protocol, and an implementation MUST NOT reset the sequence number space in this situation).

The receiver keeps track of the incoming sequence numbers on each link in a bundle and maintains the current minimum of the most recently received sequence number over all the member links in the bundle (call this M). The receiver detects the end of a packet when it receives a fragment bearing the (E)nding bit. Reassembly of the packet is complete if all sequence numbers up to that fragment have been received.

A lost fragment is detected when M advances past the sequence number of a fragment bearing an (E)nding bit of a packet which has not been completely reassembled (i.e., not all the sequence numbers between the fragment bearing the (B)eginning bit and the fragment bearing the (E)nding bit have been received). This is because of the increasing sequence number rule over the bundle. Any sequence number so detected is assumed to correspond to a fragment which has been lost.

An implementation MUST assume that if a fragment bears a (B)eginning bit, that the previously numbered fragment bore an (E)nding bit. Thus if a packet is lost bearing the (E)nding bit, and the packet whose fragment number is M contains a (B)eginning bit, the implementation MUST discard fragments for all unassembled packets through M-1, but SHOULD NOT discard the fragment bearing the new (B)eginning bit on this basis alone.

The detection of a lost fragment, whose sequence number was deduced to be U, causes the receiver to discard all fragments up to the lowest numbered fragment with an ending bit (possibly deduced) greater than or equal to U. However, the quantity M may jump into the middle of a chain of packets which can be successful completed.

RED HAT 1010
RED HAT V. CAS

Fragments may be lost due to corruption of individual packets or catastrophic loss of the link (which may occur only in one direction). This version of the multilink protocol mandates no specific procedures for the detection of failed links. The PPP link quality management facility, or the periodic issuance of LCP echo-requests could be used to achieve this.

Senders SHOULD avoid keeping any member links idle to maximize early detection of lost fragments by the receiver, since the value of M is not incremented on idle links. Senders SHOULD rotate traffic among the member links if there isn't sufficient traffic to overflow the capacity of one link to avoid idle links.

Loss of the final fragment of a transmission can cause the receiver to stall until new packets arrive. The likelihood of this may be decreased by sending a null fragment on each member link in a bundle that would otherwise become idle immediately after having transmitted a fragment bearing the (E)nding bit, where a null fragment is one consisting only of a multilink header bearing both the (B)egin and (E)nding bits (i.e., having no payload). Implementations concerned about either wasting bandwidth or per packet costs are not required to send null fragments and may elect to defer sending them until a timer expires, with the marginally increased possibility of lengthier stalls in the receiver. The receiver SHOULD implement some type of link idle timer to guard against indefinite stalls.

The increasing sequence per link rule prohibits the reallocation of fragments queued up behind a failing link to a working one, a practice which is not unusual for implementations of ISO multilink over LAPB [4].

4.2. Buffer Space Requirements

There is no amount of buffering that will guarantee correct detection of fragment loss, since an adversarial peer may withhold a fragment on one channel and send arbitrary amounts on the others. For the usual case where all channels are transmitting, you can show that there is a minimum amount below which you could not correctly detect packet loss. The amount depends on the relative delay between the channels, (D[channel-i,channel-j]), the data rate of each channel, R[c], the maximum fragment size permitted on each channel, F[c], and the total amount of buffering the transmitter has allocated amongst the channels.

When using PPP, the delay between channels could be estimated by using LCP echo request and echo reply packets. (In the case of links of different transmission rates, the round trip times should be adjusted to take this into account.) The slippage for each channel

RED HAT 1010
RED HAT V. CAS

is defined as the bandwidth times the delay for that channel relative to the channel with the longest delay, S[c] = R[c] * D[c,c-worst]. (S[c-worst] will be zero, of course!)

A situation which would exacerbate sequence number skew would be one in which there is extremely bursty traffic (almost allowing all channels to drain), and then where the transmitter would first queue up as many consecutively numbered packets on one link as it could, then queue up the next batch on a second link, and so on. Since transmitters must be able to buffer at least a maximum- sized fragment for each link (and will usually buffer up at least two) A receiver that allocates any less than S[1] + S[2] + ... + S[N] + F[1] + ... + F[N], will be at risk for incorrectly assuming packet loss, and therefore, SHOULD allocate at least twice that.

5. PPP Link Control Protocol Extensions

If reliable multilink operation is desired, PPP Reliable Transmission [6] (essentially the use of ISO LAPB) MUST be negotiated prior to the use of the Multilink Protocol on each member link.

Whether or not reliable delivery is employed over member links, an implementation MUST present a signal to the NCP's running over the multilink arrangement that a loss has occurred.

Compression may be used separately on each member link, or run over the bundle (as a logical group link). The use of multiple compression streams under the bundle (i.e., on each link separately) is indicated by running the Compression Control Protocol [5] but with an alternative PPP protocol ID.

5.1. Configuration Option Types

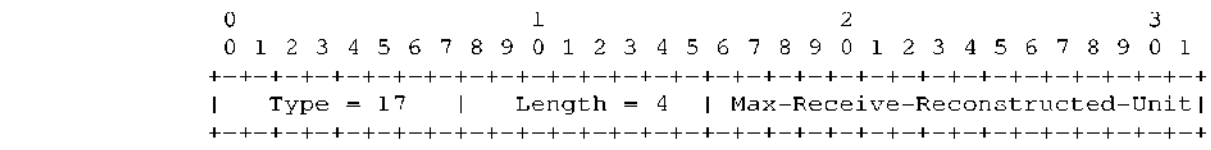
The Multilink Protocol introduces the use of additional LCP Configuration Options:

- Multilink Maximum Received Reconstructed Unit
- Multilink Short Sequence Number Header Format
- Endpoint Discriminator

RED HAT 1010
RED HAT V. CAS

5.1.1. Multilink MRRU LCP option

Figure 4: Multilink MRRU LCP option



The presence of this LCP option indicates that the system sending it implements the PPP Multilink Protocol. If not rejected, the system will construe all packets received on this link as being able to be processed by a common protocol machine with any other packets received from the same peer on any other link on which this option has been accepted.

The Max-Receive-Reconstructed unit field is two octets, and specifies the maximum number of octets in the Information fields of reassembled packets. A system MUST be able to receive the full 1500 octet Information field of any reassembled PPP packet although it MAY attempt to negotiate a smaller, or larger value. The number 1500 here comes from the specification for the MRU LCP option in PPP; if this requirement is changed in a future version of RFC 1661, the same rules will apply here.

A system MUST include the LCP MRRU option in every LCP negotiation intended to instantiate a bundle or to join an existing bundle. If the LCP MRRU option is offered on a link which is intended to join an existing bundle, a system MUST offer the same Max-Receive-Reconstruct-Unit value previously negotiated for the bundle.

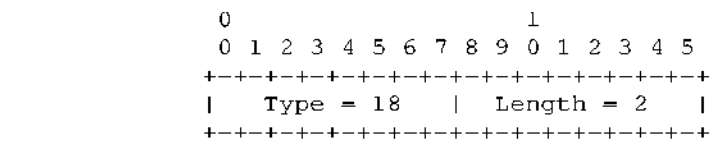
A system MUST NOT send any multilink packets on any link unless its peer has offered the MMRU LCP option and the system has configure-Ack'ed it during the most recent LCP negotiation on that link. A system MAY include the MMRU LCP option in a configure-NAK, if its peer has not offered it (until, according to PPP rules, the peer configure-Reject's it).

Note: the MRRU value conveyed im this option corresponds to the MRU of the bundle when conceptualized as a PPP entity; but the rules for the Multilink MRRU option are different from the LCP MRU option, as some value MUST be offered in every LCP negotiation, and that confirmation of this option is required prior to multilink interpretation.

RED HAT 1010
RED HAT V. CAS

5.1.2. Short Sequence Number Header Format Option

Figure 5: Short Sequence Number Header Format Option

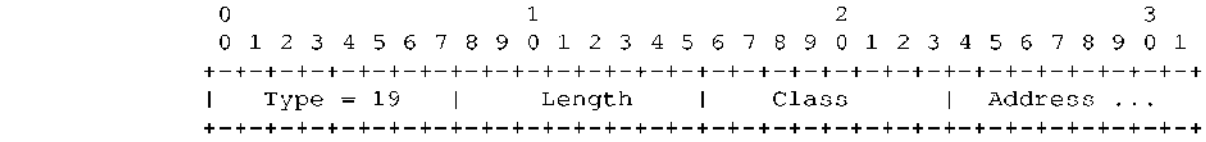


This option advises the peer that the implementation wishes to receive fragments with short, 12 bit sequence numbers. When a peer system configure-Ack's this option, it MUST transmit all multilink packets on all links of the bundle with 12 bit sequence numbers or configure-Reject the option. If 12 bit sequence numbers are desired, this option MUST be negotiated when the bundle is instantiated, and MUST be explicitly included in every LCP configure request offered by a system when the system intends to include that link in an existing bundle using 12 bit sequence numbers. If this option is never negotiated during the life of a bundle, sequence numbers are 24 bits long.

An implementation wishing to transmit multilink fragments with short sequence numbers MAY include the multilink short sequence number in a configure-NAK to ask that the peer respond with a request to receive short sequence numbers. The peer is not compelled to respond with the option.

5.1.3. Endpoint Discriminator Option

Figure 7: Endpoint Discriminator Option



The Endpoint Discriminator Option represents identification of the system transmitting the packet. This option advises a system that the peer on this link could be the same as the peer on another existing link. If the option distinguishes this peer from all others, a new bundle MUST be established from the link being negotiated. If this option matches the class and address of some other peer of an existing link, the new link MUST be joined to the bundle containing the link to the matching peer or MUST establish a new bundle, depending on the decision tree shown in (1) through (4) below.

RED HAT 1010
RED HAT V. CAS

To securely join an existing bundle, a PPP authentication protocol [3] must be used to obtain authenticated information from the peer to prevent a hostile peer from joining an existing bundle by presenting a falsified discriminator option.

This option is not required for multilink operation. If a system does not receive the Multilink MRRU option, but does receive the Endpoint Discriminator Option, and there is no manual configuration providing outside information, the implementation MUST NOT assume that multilink operation is being requested on this basis alone.

As there is also no requirement for authentication, there are four sets of scenarios:

(1) No authentication, no discriminator:
All new links MUST be joined to one bundle, unless there is manual configuration to the contrary. It is also permissible to have more than one manually configured bundle connecting two given systems.

(2) Discriminator, no authentication:
Discriminator match -> MUST join matching bundle, discriminator mismatch -> MUST establish new bundle.

(3) No discriminator, authentication:
Authenticated match -> MUST join matching bundle, authenticated mismatch -> MUST establish new bundle.

(4) Discriminator, authentication:
Discriminator match and authenticated match -> MUST join bundle, discriminator mismatch -> MUST establish new bundle, authenticated mismatch -> MUST establish new bundle.

The option contains a Class which selects an identifier address space and an Address which selects a unique identifier within the class address space.

This identifier is expected to refer to the mechanical equipment associated with the transmitting system. For some classes, uniqueness of the identifier is global and is not bounded by the scope of a particular administrative domain. Within each class, uniqueness of address values is controlled by a class dependent policy for assigning values.

Each endpoint may chose an identifier class without restriction. Since the objective is to detect mismatches between endpoints erroneously assumed to be alike, mismatch on class alone is sufficient. Although no one class is recommended, classes which have

RED HAT 1010
RED HAT V. CAS

universally unique values are preferred.

This option is not required to be supported either by the system or the peer. If the option is not present in a Configure-Request, the system MUST NOT generate a Configure-Nak of this option for any reason; instead it SHOULD behave as if it had received the option with Class = 0, Address = 0. If a system receives a Configure-Nak or Configure-Reject of this option, it MUST remove it from any additional Configure-Request.

The size is determined from the Length field of the element. For some classes, the length is fixed, for others the length is variable. The option is invalid if the Length field indicates a size below the minimum for the class.

An implementation MAY use the Endpoint Discriminator to locate administration or authentication records in a local database. Such use of this option is incidental to its purpose and is deprecated when a PPP Authentication protocol [3] can be used instead. Since some classes permit the peer to generate random or locally assigned address values, use of this option as a database key requires prior agreement between peer administrators.

The specification of the subfields are:

Type
19 = for Endpoint Discriminator

Length
3 + length of Address

Class
The Class field is one octet and indicates the identifier address space. The most up-to-date values of the LCP Endpoint Discriminator Class field are specified in the most recent "Assigned Numbers" RFC [7]. Current values are assigned as follows:

0 Null Class

1 Locally Assigned Address

2 Internet Protocol (IP) Address

3 IEEE 802.1 Globally Assigned MAC Address

4 PPP Magic-Number Block

RED HAT 1010
RED HAT V. CAS

5 Public Switched Network Directory Number

Address

The Address field is one or more octets and indicates the identifier address within the selected class. The length and content depend on the value of the Class as follows:

Class 0 - Null Class

Maximum Length: 0

Content:
This class is the default value if the option is not present in a received Configure-Request.

Class 1 - Locally Assigned Address

Maximum Length: 20

Content:

This class is defined to permit a local assignment in the case where use of one of the globally unique classes is not possible. Use of a device serial number is suggested. The use of this class is deprecated since uniqueness is not guaranteed.

Class 2 - Internet Protocol (IP) Address

Fixed Length: 4

Content:

An address in this class contains an IP host address as defined in [8].

Class 3 - IEEE 802.1 Globally Assigned MAC Address

Fixed Length: 6

Content:

An address in this class contains an IEEE 802.1 MAC address in canonical (802.3) format [9]. The address MUST have the global/local assignment bit clear and MUST have the multicast/specific bit clear. Locally assigned MAC addresses should be represented using Class 1.

RED HAT 1010
RED HAT V. CAS

Class 4 - PPP Magic-Number Block

Maximum Length: 20

Content:

This is not an address but a block of 1 to 5 concatenated 32 bit PPP Magic-Numbers as defined in [2]. This class provides for automatic generation of a value likely but not guaranteed to be unique. The same block MUST be used by an endpoint continuously during any period in which at least one link is in the LCP Open state. The use of this class is deprecated.

Note that PPP Magic-Numbers are used in [2] to detect unexpected loopbacks of a link from an endpoint to itself. There is a small probability that two distinct endpoints will generate matching magic-numbers. This probability is geometrically reduced when the LCP negotiation is repeated in search of the desired mismatch, if a peer can generate uncorrelated magic-numbers.

As used here, magic-numbers are used to determine if two links are in fact from the same peer endpoint or from two distinct endpoints. The numbers always match when there is one endpoint. There is a small probability that the numbers will match even if there are two endpoints. To achieve the same confidence that there is not a false match as for LCP loopback detection, several uncorrelated magic-numbers can be combined in one block.

Class 5 - Public Switched Network Directory Number

Maximum Length: 15

Content:

An address in this class contains an octet sequence as defined by I.331 (E.164) representing an international telephone directory number suitable for use to access the endpoint via the public switched telephone network [10].

6. Initiating use of Multilink Headers

When the use of the Multilink protocol has been negotiated on a link (say Y), and the link is being added to a bundle which currently contains a single existing link (say X), a system MUST transmit a Multilink-encapsulated packet on X before transmitting any Multilink-

RED HAT 1010
RED HAT V. CAS

encapsulated packets on Y.

Since links may be added and removed from a bundle without destroying the state associated with it, the fragment should be assigned the appropriate (next) fragment number. As noted earlier, the first fragment transmitted in the life of a bundle is assigned fragment number 0.

7. Closing Member links

Member links may be terminated according to normal PPP LCP procedures using LCP Terminate-Request and Terminate-Ack packets on that member link. Since it is assumed that member links usually do not reorder packets, receipt of a terminate ack is sufficient to assume that any multilink protocol packets ahead of it are at no special risk of loss.

Receipt of an LCP Terminate-Request on one link does not conclude the procedure on the remaining links.

So long as any member links in the bundle are active, the PPP state for the bundle persists as a separate entity. However, if the there is a unique link in the bundle, and all the other links were closed gracefully (with Terminate-Ack), an implementation MAY cease using multilink
headers.

If the multilink procedure is used in conjunction with PPP reliable transmission, and a member link is not closed gracefully, the implementation should expect to receive packets which violate the increasing sequence number rule.

8. Interaction with Other Protocols

In the common case, LCP, and the Authentication Control Protocol would be negotiated over each member link. The Network Protocols themselves and associated control exchanges would normally have been conducted once, on the bundle.

In some instances it may be desirable for some Network Protocols to be exempted from sequencing requirements, and if the MRU sizes of the link did not cause fragmentation, those protocols could be sent directly over the member links.

Although explicitly discouraged above, if there were several member links connecting two implementations, and independent sequencing of two protocol sets were desired, but blocking of one by the other was not, one could describe two multilink procedures by assigning

RED HAT 1010
RED HAT V. CAS

multiple endpoint identifiers to a given system. Each member link, however, would only belong to one bundle. One could think of a physical router as housing two logically separate implementations, each of which is independently configured.

A simpler solution would be to have one link refuse to join the bundle, by sending a Configure-Reject in response to the Multilink LCP option.

9. Security Considerations

Operation of this protocol is no more and no less secure than operation of the PPP authentication protocols [3]. The reader is directed there for further discussion.

10. References

[1] Leifer, D., Sheldon, S., and B. Gorsline, "A Subnetwork Control Protocol for ISDN Circuit-Switching", University of Michigan (unpublished), March 1991.

[2] Simpson, W., Editor, "The Point-to-Point Protocol (PPP)", STD 51, RFC 1661, Daydreamer, July 1994.

[3] Lloyd, B., and W. Simpson, "PPP Authentication Protocols", RFC 1334, Lloyd Internetworking, Daydreamer, October 1992.

[4] International Organisation for Standardization, "HDLC - Description of the X.25 LAPB-Compatible DTE Data Link Procedures", International Standard 7776, 1988

[5] Rand, D., "The PPP Compression Control Protocol (CCP)", PPP Extensions Working Group, RFC 1962, June 1996.

[6] Rand, D., "PPP Reliable Transmission", RFC 1663, Novell, July 1994

[7] Reynolds, J., and J. Postel, "Assigned Numbers", STD 2, RFC 1700, USC/Information Sciences Institute, October 1994.

[8] Postel, J., Editor, "Internet Protocol - DARPA Internet Program Protocol Specification", STD 5, RFC 791, USC/Information Sciences Institute, September 1981.

[9] Institute of Electrical and Electronics Engineers, Inc., "IEEE Local and Metropolitan Area Networks: Overview and Architecture", IEEE Std. 802-1990, 1990.

RED HAT 1010
RED HAT V. CAS

[10] The International Telegraph and Telephone Consultative Committee (CCITT), "Numbering Plan for the ISDN Area", Recommendation I.331 (E.164), 1988.

[11] Simpson, W., Editor, "PPP LCP Extensions", RFC 1570, Daydreamer, January 1994.

11. Differences from RFC 1717

This section documents differences from RFC 1717. There are restrictions placed on implementations that were absent in RFC 1717; systems obeying these restrictions are fully interoperable with RFC 1717 - compliant systems.

11.1. Negotiating Multilink, per se

RFC 1717 permitted either the use of the Short Sequence Number Header Format (SSNHF) or the Maximum Reconstructed Receive Unit (MRRU) options by themselves to indicate the intent to negotiate multilink. This specification forbids the use of the SSNHF option by itself; but does permit the specific of both options together. Any implementation which otherwise conforms to rfc1717 and also obeys this restriction will interoperate with any RFC 1717 implementation.

11.2. Initial Sequence Number defined

This specification requires that the first sequence number transmitted after the virtual link has reached to open state be 0.

11.3. Default Value of the MRRU

This specfication removes the default value for the MRRU, (since it must always be negotiated with some value), and specifies that an implementation must be support an MRRU with same value as the default MRU size for PPP.

11.4. Config-Nak of EID prohibited

This specification forbids the config-Naking of an EID for any reason.

11.5. Uniformity of Sequence Space

This specification requires that the same sequence format be employed on all links in a bundle.

RED HAT 1010
RED HAT V. CAS

11.6. Commencing and Abating use of Multilink Headers

This memo specifies how one should start the use of Multilink Headers when a link is added, and under what circumstances it is safe to discontinue their use.

11.7. Manual Configuration and Bundle Assignment

The document explicitly permits multiple bundles to be manually configured in the absence of both the Endpoint Descriminator and any form of authentication.

RED HAT 1010
RED HAT V. CAS

13. Authors' Addresses

Keith Sklower
Computer Science Department
384 Soda Hall, Mail Stop 1776
University of California
Berkeley, CA 94720-1776

Phone: (510) 642-9587
EMail: sklower@CS.Berkeley.EDU

Brian Lloyd
Lloyd Internetworking
3031 Alhambra Drive
Cameron Park, CA 95682

Phone: (916) 676-1147
EMail: brian@lloyd.com

Glenn McGregor
Lloyd Internetworking
3031 Alhambra Drive
Cameron Park, CA 95682

Phone: (916) 676-1147
EMail: glenn@lloyd.com

Dave Carr
Newbridge Networks Corporation
600 March Road
P.O. Box 13600
Kanata, Ontario,
Canada, K2K 2E6

Phone: (613) 591-3600
EMail: dcarr@Newbridge.COM

Tom Coradetti
Sidewalk Software
1190 Josephine Road
Roseville, MN 55113

Phone: (612) 490 7856
EMail: 70761.1664@compuserve.com

RED HAT 1010
RED HAT V. CAS

Multiple Paths Protocol for a Cluster Type Network

Hiroshi Mineno, Lu Junshu, Ken Ohta, Masahiro Aono, Tetsuo Ideguchi† and Tadanori Mizuno

†Faculty of Information Science and Technology
Aichi Prefectural University
Nagakute, aichi, 480-1198 Japan

Graduate School of Science and Engeering
Shizuoka University
3-5-1 Johoku, Hamamatsu, 432 Japan

Abstract

We can access Internet by carrying a portable computer and using the wireless communication. The wireless network with PHS(Personal Handy phone System) and portable cellular telephone has only rates of tens of Kbps to a few Mbps. Compared with the cable network, the transfer rate cannot generally satisfy a highly developed communication services such as large file transfer and real-time communications. This paper proposes a protocol, SHAKE, for sharing multiple paths in cluster type network that is a kind of LAN in which some mobile hosts temporarily connect mutually. SHAKE provides the functions for composing cluster type network, and dispersing traffic efficiently by measuring transfer rate and round-trip time. As a mobile host has only low transfer capacity in individual to communicate with outside, if whole capacities of other hosts which compose cluster type network are shared, we can get larger transfer capacity and satisfy the required communication services.

1 Introduction

With the rapid development of the network technology and the improvement of the computer's performance, we can access Internet by carrying a portable computer and using the wireless communication. In the present status, the wireless network with PHS(Personal Handy phone System), portable cellular telephone or in-house wireless LAN has only rates of tens of Kbps to a few Mbps. Compared with the wired network, the transfer rate cannot generally satisfy a highly developed communication services such as transfer of large files or multimedia data, and real-time communication which should guarantee the delay time. This means that the demands for capacity may sometimes exceed what is available, and packets may be lost.

In this paper, we propose a protocol SHAKE (SHAred multiple paths protocol for a cluster type networK Environment), for sharing multiple paths in a cluster type network. Cluster type network is a kind of LAN in which some mobile hosts temporarily connect mutually. In mobile units such as trains, ships and aircrafts where some persons gather, there are two ways to get larger capacity by composing the cluster type network. One way is that the host which has larger or more stable path is made to represent and communicate with outside. No mobile host which connects cluster type network needs to have its own outside path. The other is that as a mobile host has only low transfer capacity in individual to communicate with outside, if whole capacities of other hosts which compose cluster network are shared, we can get larger transfer capacity and spread burst traffic equally.

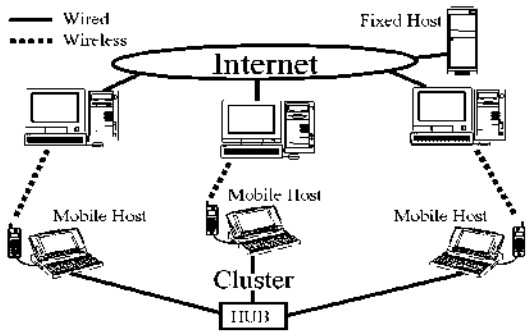


Figure 1: An Example of the System.

We focus the latter(Fig. 1). We can select several kinds of paths to communicate with outside. So the communication media such as PHS or portable cellular telephone can be freely selected by each mobile host. But using different paths to communicate, the quality of each paths such as transfer rate, delay time, and the rate of packet loss are different, and for this reason, it is important to consider how to distribute the burst traffic and how to reconstruct the packets.

This paper is organized as follows. The next section describes the background of this research work

RED HAT 1010
RED HAT V. CAS

and the related research works; Section 3 describes the shared multiple paths protocol, SHAKE and its functions; Section 4 talks about the implementation of the prototype; And Section 5 concludes this paper.

2 Related Works

2.1 Traffic Dispersion

Generally, aggregating the resource can improve the performance and efficiency in shared system or communication network[1]. One of the resource aggregation methods is that the traffic is transmitted in parallel over multiple paths.

Maxemchuk presented that for load balancing and fault handling in packet-switched networks, it is better in space rather than in time to disperse the traffic[2, 3, 4]. According to this method, a message from a source is distributed into several submessages, which are transmitted in parallel over different paths in network. And as a transmission error on one path is independent from errors on the other paths, forward error correcting codes can be successfully used. Furthermore, in an algorithm proposed by Lee and Liew[5], the traffic from a source is partitioned into submessages, each consisting of K packets. This algorithm encodes the K packets into N>K packets, which are transmitted in parallel over N separate paths. When any K of the N packets are correctly received, the original message can be reconstructed. And if the bit error detection is used for every packet, the algorithm is capable of correcting N−K packet errors in each submessage.

IETF standardizes MP(Multilink Point-to-Point Protocol) with the technique which consists of a number of different switched WAN services such as ISDN and ATM to form a virtual path[6, 7]. It can strengthen the functions of PPP(Point-to-Point Protocol) data communications between two points, and a number of virtual connections can be set up between equipments by taking advantages of characters of switched WAN service. If necessary, the bandwidth which a user requested can be divided, and a number of physical lines and virtual lines can be used as a single logical line.

The technology of aggregating multiple communication paths and dispersing traffic into each communication path are being developed to improve the performance and efficiency of transmission. And they are generally used to communicate with different kinds of WAN lines which are aggregated among routers.

2.2 Mobile Distributed Co-operation System

In mobile computing environment, a mobile host works as a terminal for personal operation. Furthermore it is also considered to form a group to co-operate with. There is a mobile groupware which supports co-operation for a number of mobile hosts. In the area where a number of mobile hosts can communicate mutually, the number of mobile hosts can construct a temporary network to support co-operation for people gathered in the area. The network is called cluster type network or ad hoc network, and some research works are being developed [8, 9, 10].

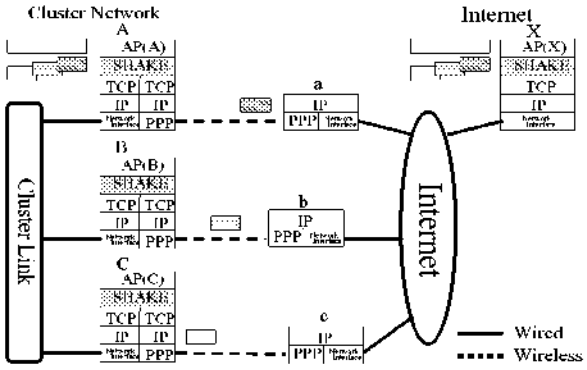


Figure 2: SHAKE Protocol Stack.

In addition, it was also considered how to communicate with outside networks. The one is that in some kinds of transportations such as trains, ships or aircrafts, if there is a communication system with high communication capacity which works as a representative of all mobile hosts to communicate with outside networks, the transfer time and cost can be reduced. And the other is that as a mobile host has only low transfer rate in individual to communicate with outside networks, but if whole capacities of other hosts which compose cluster network are shared, we can get larger transfer capacity.

So the cluster has some advantages which a single mobile host hasn't. And the latter communication protocol in which other hosts' communication capacity can be shared is called SHAKE, which we are developing. By using this, the communication paths of other mobile hosts which compose the cluster are shared and they are considered logically one path.

3 Shared Multiple Paths

3.1 SHAKE

We think there are severaral advantages of the sharing multiple paths: The end-to-end transfer rate of data can be increased and the delay time be reduced. The bursty traffic is dispersed into several paths and network load is equalized. Transmission errors on each path are independent from errors on the other paths.

RED HAT 1010
RED HAT V. CAS

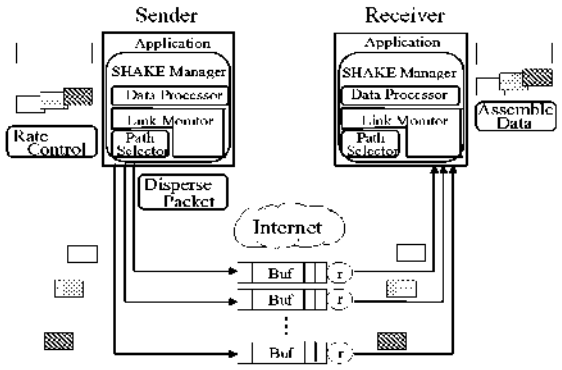


Figure 3: Illustration of SHAKE Modules.

To transfer packets redundantly, the characteristics of delay time and packet loss on different paths can be reduced, and reliability can be improved. It can form a cluster in which each mobile host can gather anywhere. Any mobile host in a cluster can freely select any kinds of medium for communication between the cluster and the outside network. If one of mobile hosts in cluster has not a path to communicate with outside, it can transfer data by relaying other's path.

We are now developing the protocol SHAKE to realize the shared multiple paths. SHAKE is implemented at all mobile hosts which form a cluster (cluster means cluster type network) and at fixed hosts which are connected to Internet and communicate with one of mobile hosts in a cluster (Fig. 2). And for easy to implement SHAKE, it is built between the Transport Layer and the Application Layer. So the mobile host which isn't implemented SHAKE can't share multiple paths.

In the cluster, we call a mobile host that has a communication path with outside network (like Internet) as a relay host. Each relay host in the cluster has two network interfaces: one is used to communicate with outside network, and the other is assigned the local private IP address before forming a cluster. Cluster is formed by using Ethernet, serial cable, wireless LAN, IrDA and so on, and the shape of the network can be considered as ring type network or star type network. A mobile host can freely select any communication path and any kind of medium to communicate with outside. For those hosts which have no connection with outside, the communication can be also done through the relay hosts. The logical link which connects each mobile host to form the cluster is called "cluster link". A mobile host became a part of the cluster if it is connected by cluster link.

3.2 Functions of SHAKE

As a protocol of Transport Layer, SHAKE uses TCP to transmit data. TCP connection is formed between mobile hosts in the cluster and fixed hosts in the Internet. There are three kinds of SHAKE packets which go through this connection, each is identified by special SHAKE header; the first is data packet which contains data from the Application Layer; the second is control packet to form a cluster, and the last is report packet which is used to report communication status. The protocol SHAKE is realized from the following 4 modules.

- **Link monitoring module**:
- **Path selecting module**:
- **Data processing module**:
- **SHAKE management module**:

The relationship among these modules is shown in Fig.3. From these modules, we can realize the SHAKE for sharing multiple paths in cluster type network.

3.2.1 Link Monitoring Module

This module monitors the status of communication paths with outside networks and estimate the sending rate. Here the status of communication paths means the theoretical bandwidth of used medium, throughput which changes actively, transit delay, cut of communication paths, etc.. And if a notable change of the status is happened, it changes the sending rate to use each path's bandwidth effectively.

Here we will talk about the estimation of sending rate for transferring data. The method is to control the definite amount of data which pass the network. We consider the network as a large buffer (we call it a network buffer) which has it's own queue and sending rate. For using each path effectively, it is better to leave network buffer in constant value to avoid that the receiving buffer in the object is often empty. But the problem will be happened if too much data is sent, which can cause the transit delay largely, so it is important to control the amount of data in the network buffer. Like in Fig. 3, it seems that each communication path has a buffer and all buffers are managed by Link Monitoring module. The SHAKE evaluates how many packets have been sent to the object and how many packets still overstock in each network buffer. The SHAKE can estimate sending rate to keep amount of data in all network buffers.

The remained amount of data in network buffers can be evaluated as follows. In sending host, each of the dispersed packets is added a sending time stamp

RED HAT 1010
RED HAT V. CAS

ST_1 in SHAKE header and sent to the receiving host. On the other side, at regular intervals, the receiving host sends back a receiving report to the sending host, which contains the newest received packet's ST_1, RT_1 which is the time it received, RT_2 which is the time this report was sent, and R_{recv} which is rate of receiving. So when the sending host received the report at the time ST_2, the sending host can estimate the RTT by the following expression. And to avoid the wide fluctuation of RTT_{avg} which means average of RTT, it is estimated as follows by using low-pass filter $(0 < \alpha \leq 1)$.

$$RTT = (ST_2 - ST_1) - (RT_2 - RT_1) \quad (1)$$

$$RTT_{avg} = (1 - \alpha)RTT_{avg} + \alpha RTT \quad (2)$$

And then Buf_{cur} which means current network buffer is estimated as follows by using the R_{recv}, RTT_{avg}, and RTT_{min} which is minimum RTT value till the time:

$$Buf_{cur} = R_{recv}(RTT_{avg} - RTT_{min}) \quad (3)$$

Comparing the expected value Buf_{des} with current amount of data in network buffer Buf_{cur}, the sending rate R_{snd} is increased if expected value is small, or reduced if expected value is large. And using the receiving rate R_{rcv} in the receiving report, we can calculate the accurate sending rate:

$$R_{snd} = R_{recv} + \frac{(Buf_{des} - Buf_{cur})}{interval} \quad (4)$$

3.2.2 Path Selecting Module

A sending host gets information about communication paths with outside from the link monitoring module, and then distributes the data packets which come from the application layer. If each communication path has same length, type, throughput and transit delay, the packets can be distributed equally to each of communication paths in order. But actually, in each path, characters of network are different since each path can be selected freely by mobile hosts. So the distribution rate of each path must be changed according to the status of each communication path. This is done by analyzing the information about the sending rate R_{snd}, and disconnection report from link monitoring module.

The distribution procedure is to distribute packets according to the rate of each path's R_{snd}, and adjust the amount of data in each network buffer. If there are n paths in cluster and the sending rate is R_i in path $i(i \leq n)$, the packet distribution rate P_i of path i can be calculated as follows. And according to the packet distribution rate P_i, the packets can be distributed for each communication path.

$$P_i = \frac{R_i}{\sum_{k=1}^{n} R_k} \quad (5)$$

3.2.3 Data Processing Module

The received data from application layer are divided into a number of packets. Each packet is transmitted to transport layer after it is added a special header of SHAKE.This packet is a data packet, and the header includes: packet identifier (Data), link number, packet length, sequence number, ST1, and data. And there is a report packet which includes: packet identifier (Report), link number, packet length, ST1, RT1, RT2, and receiving rate of the link. Fig. 4, Fig. 5 show the format of each SHAKE packet. And last, there is a control packet which is only used to form a cluster and make a link table.

When each host receives a SHAKE packet, it identifies the packet whether it's a data packet or a report packet or a control packet by checking the identifier in the header, and according to the link number in the header, it judges where to relay the packet. Each host has a table which relates host with link number. So if the packet is destined the host itself, the packet is given to the application and the others are relayed to the next host.

Here we explain how to put in order the received data packets with irregular sequence since they were transmitted from multiple paths. Although the path selecting module selects paths from which packets be received by destination in right sequence, actually it's difficult to do it completely because of different characters like transit delay and length of each path. If a packet with irregular sequence is received by SHAKE, it cannot be given to the application directly. SHAKE will store it to the buffer temporarily, and then if any packet which can be given to the application is received, SHAKE will check the buffer if there is any packet with the following sequence number. If there is, of course it will be given to the application, and SHAKE will continue to search whether there is any packet else in the buffer. If no one else, SHAKE receives the next packet and repeats the above processing.

3.2.4 SHAKE Management Module

The work of this module includes controlling the other modules, forming the cluster, ensuring communication paths and so on. The application of the upper layer

RED HAT 1010
RED HAT V. CAS

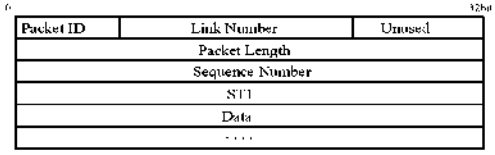


Figure 4: Data Packet Format

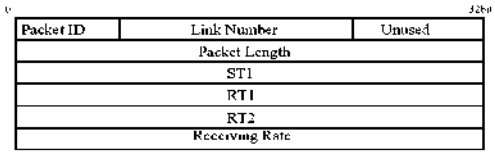


Figure 5: Report Packet Format.

needn't to consider how and on which paths are selected to transmit data. It realizes the shared multilink procedure by using multimedia through multiple paths to form a logical path.

4 Implementation

In this section, we present the implementation environment and the evaluation of the prototype which we are developing now.

4.1 The Implementation Environment

The cluster network is formed by mobile hosts (Toshiba SatellitePro420, CPU: Pentium100MIIz Memory:16.0MB, and Mitsubishi Amity SP, CPU: 486DX4-75MIIz Memory:16.0MB) which installed Windows95 and with two PCMCIA card slots. For sending packets within the cluster, the network interface which is connected to the cluster is assigned the private IP address like 192.168.X.X, and each host connected with about 1m's 10Base-T and 4 port HUB to form a cluster. The other network interface is assigned a normal IP address to connect the Internet, which may be assigned by DHCP or decided value.

For the implementation of the SHAKE prototype, we developed a simply special data transfer application with Visual C++, and the SHAKE is used in the communication section of the application. It is implemented at all mobile hosts which form a cluster, and the host in the Internet which communicates with one of the host in cluster network. Then it can divide an application data into a number of determined size SHAKE packets and distributes to the destination. And in destination, the received packets is reconstructed and given to the application in order.

4.2 The Status of Prototype

The present link monitoring module forms a path information table of all end-to-end path information when the cluster is formed, and the information is updated by receiving the report packets. When some new hosts are added to the cluster, new sections are added to this table after new connecting. It performs a role as a routing table indicated among link number and relay host and destination host, and it contains various information like theoretical bandwidth, round-trip delay, and receiving rate. These values are estimated by receiving report packet from the recieving host at regular intervals, and each link's suitable sending rate is also estimated by this report.

The path selecting module searches if the destination host is in the information table when the application requests for sending data. And then it checks the path, and simply sends through these paths in order. That is to say, the function which can change the packet distribution rate according to the each path's sending rate is being developed.

The data processing module is almost completely implemented. It divides application data into a number of SHAKE packets, and adds a special SHAKE header to each packet. When the packet is received, it is processed to determine if it is relayed or be given to the application according to the information of header. If the packet is to be given to the application, it should be put back in order since they may be transmitted in different paths and received in irregular sequence. In the present method, the sequence number in SHAKE header will be checked to determine the action. If the packet can't be given to the application, it is stored in buffer and its sequence number will be used as a key for searching. When a new packet is received and given to the application, the SHAKE checks if the next packet is stored in buffer. If it is, the packet will be took out and haded to the application, then the receiver repeats the above processing.

The tasks of SHAKE management module are checking among hosts and managing any other modules. It keeps communications with the application directly, and an additional window of SHAKE is opened for the request of forming a cluster.

4.3 Experiments

In our experiments, we used two mobile hosts to form a cluster network, and used wireless LAN links (2.4GHz frequency hopping spread spectrum) for communication with outside networks. We did the following three kinds of experiments.

1. In experiment one, we used one access point for wireless LAN, and the mobile host communicates with the fixed host which connected with our laboratory's LAN by using a wireless LAN (Fig.6).

RED HAT 1010
RED HAT V. CAS



Figure 6: Experiment 1.

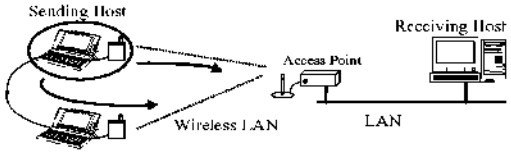


Figure 7: Experiment 2.

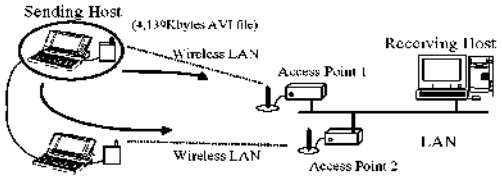


Figure 8: Experiment 3.

2. In experiment two, we used one access point for wireless LAN, and two wireless LAN adapters for sharing the paths. The cluster is formed by two mobile hosts, only one mobile host communicates with the fixed host which connected with the laboratory's LAN (Fig. 7).

3. In experiment three, we used two access points. We set up two channels for them to share the communication paths under the situation that there is no affection between them. The cluster is formed by two mobile hosts, and done the same experiment (Fig. 8).

The used file is about 1Mbytes (1139kbytes) AVI file. It was sent from the client host in the cluster to server host in fixed network. Then in the server host, we investigated the amount of received data and the receiving time of data (which means the time beginning the first data packet received to the time that all data packets were received). And the throuput is estimated at the server host. The relationship is shown in graphs (Fig. 9).

Next we did above experiments with PHS terminals on PIAFS mode. PIAFS (the abbreviation for PHS Internet Access Forum Standards) is the data transmission standards to operate a high speed data communication service with 32Kbps access via PHS. This environment is shown in Fig. 10.

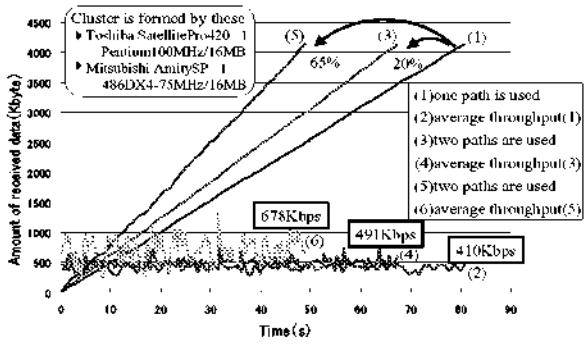


Figure 9: Results of Wireless LAN experiments.

4.4 Results

There are six graphs in the Figure 9. Graph (1)(3)(5) show the relationship between the received time and the amount of received data, and graph (2)(4)(6) show the average of the throughput which is calculated when five data packets are received. Graph (1)(2) are the results of experiment 1, graph (3)(4) are experiment 2, and graph (5)(6) are experiment 3.

The results shows that to form a cluster network and share multiple paths can reduce the transfer time and improve the throughput comparing with using a single wireless LAN adapter. When one wireless path is used, the average of throughput is about 410Kbps, and when two wireless paths are used with same access point, the average is about 491Kbps. In spite of sharing the same channel with two wireless LAN adapters, the average of throughput is improved a little (20%). And if the using channels are differ from, the throughput is better (65%).

The communication speed is changed according to different conditions, but the average of throughput which is guaranteed by the manufacture company is about 400 Kbps. So both of throughputs we calculated from graphs (2)(4) of Fig. 9 are the results we expected. And we know that if sharing the same channel, we can use the channel effectively and the average of throughput nearly approaches the limit value of using one channel.

From the results of experiment 3, graph (5)(6), we could get the 65% higher throughput than using one wireless path. The value is not yet twofold, but if the prototype is more developed, we may improve the throughput better. And in graph (6), the average of throughput is unstable. We think that the reason why the graph is so unstable is the function of resequencing packets. Some functions, such as the sending rate control and the distribution rate control, were not im-

RED HAT 1010
RED HAT V. CAS

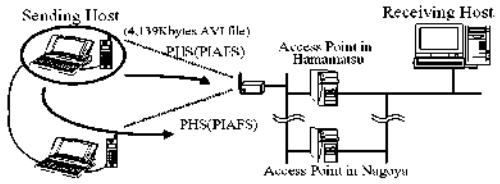


Figure 10: Experiment with PHS terminals.

plemented in the present prototype, so many packets are received out of order. After we implement the functions, the performance could be better.

Next we will talk the results of the experiments with PHS terminals on PIAFS mode. The results are shown in the graph (Fig. 11). Graph (1)(2) are the results of experiment 1 via PIAFS access point in Nagoya, and the average of throughput is about 19Kbps. And graph (3)(4) are the results of experiment 1 via PIAFS access point in Hamamatsu, and the average of throughput is about 26Kbps. Nagoya is about 100kms away from Hamamatsu, so the defference of delay between them was about 80ms (calculated from the measured RTT). Graph (5)(6) are the results of using two PIAFS links. Compared with the case of using one PIAFS link, the average of throughput is improved 93% better than the sum of the former two cases.

5 Conclusions

We proposed a protocol, SHAKE, for sharing multiple paths in a cluster type network in which a number of mobile hosts temporarily connect mutually. Each host in the cluster can share other hosts' communication links. And SHAKE consists of four parts: link monitoring module, path selecting module, data processing module and SHAKE management module. Till now we implemented the prototype of SHAKE and did communication tests with wireless LAN systems and PHS terminals.

From now on, we plan to continue the implementation of the protocol, to simulate the affection of a cluster's size, to evaluate the efficiency of control of transmission rate and data transmission with distributed packets, and so on.

Acknowledgments

The authors gratefully acknowledge the help of Junshu Lu who assisted us to write in English, and Nicholas F. Maxemchuk who supported several references.

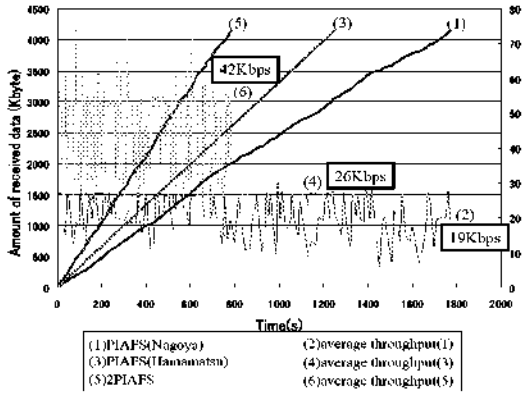


Figure 11: Results of PIIS experiments.

References

[1] E. Gustafsson, and G. Karlsson, "A Literature Survey on Traffic Dispersion," *IEEE Network March/April*, Vol. 11, No. 2, pp. 28–36, 1997.

[2] N. F. Maxemchuk, "Dispersity Routing", *ICC'75*, pp. 41–10–41–12, June 1975.

[3] N. F. Maxemchuk, "Dispersity Routing in High-Speed Networks", *Computer Networks and ISDN Systems*, Vol. 25, No. 6, pp. 645–661, January 1993.

[4] N. F. Maxemchuk, "Dispersity Routing in ATM Network", *IEEE INFOCOM'93*, pp. 347 357, March 1993.

[5] T. T. Lee and S. C. Liew, "Parallel Communications for ATM Network Control and Management", *IEEE GLOBECOM'93*, Vol. 1, pp. 442 446, December 1993.

[6] K. Sklower, B. Lloyd, G. McGregor, and D. Carr, "The PPP Multilink Protocol (MP)", November 1994. RFC1717.

[7] K. Sklower, B. Lloyd, G. McGregor, D. Carr and T. Coradetti, "The PPP Multilink Protocol (MP)", August 1996. RFC1990.

[8] P. Gupta and P. R. Kumar, "A system and traffic dependent adaptive routing algorithm for ad hoc networks", in Proc. IEEE 36th Conf. on Decision and Control, pp. 2375 2380, 1997.

[9] B. Das, R. Sivakumar, and V. Bharghavan, "Routing in Ad-Hoc Networks Using a Spine", in Proc. 6th International Conference on Computer Communications and Networks, September, 1997.

[10] "Mobile Ad Hoc Networking (MANet)", http://tonnant.itd.nrl.navy.mil/manet/manet_home.html.

RED HAT 1010
RED HAT V. CAS

Electronic Patent Application Fee Transmittal

Application Number:	14512414
Filing Date:	11-Oct-2014
Title of Invention:	MULTIPATH COMMUNICATION DEVICES AND METHODS
First Named Inventor/Applicant Name:	Eric M. Delangis
Filer:	Harold C. McGurk
Attorney Docket Number:	ED 2000

Filed as Small Entity

Filing Fees for Utility under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Claims in excess of 20	2202	3	40	120
Independent Claims in Excess of 3	2201	1	210	210
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				

RED HAT 1010
RED HAT V. CAS

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				
Miscellaneous:				
Submission- Information Disclosure Stmt	2806	1	90	90
		Total in USD ($)		**420**

Electronic Acknowledgement Receipt	
EFS ID:	24127054
Application Number:	14512414
International Application Number:	
Confirmation Number:	9530
Title of Invention:	MULTIPATH COMMUNICATION DEVICES AND METHODS
First Named Inventor/Applicant Name:	Eric M. Delangis
Customer Number:	64828
Filer:	Harold C. McGurk
Filer Authorized By:	
Attorney Docket Number:	ED 2000
Receipt Date:	18-NOV-2015
Filing Date:	11-OCT-2014
Time Stamp:	20:06:27
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	Credit Card
Payment was successfully received in RAM	$420
RAM confirmation Number	17672
Deposit Account	
Authorized User	
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:	

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Amendment/Req. Reconsideration-After Non-Final Reject	14512414_Amendment_11-18-2015.pdf	121852 877a24effce0a61a3c0ddbc6e9624cb03ec4 7de5	no	10
Warnings:					
Information:					
2	Information Disclosure Statement (IDS) Form (SB08)	14512414_IDS_11-18-2015.pdf	612322 676db3a983620d312a850faacb96c59c935 16d8b	no	4
Warnings:					
Information:					
3	Transmittal Letter	14512414_IDS_Transmittal_Letter.pdf	72505 995cb139a4f91fcbc30240c94237764401b9 d52c	no	1
Warnings:					
Information:					
4	Non Patent Literature	rfc1990_Multilink_PPP.pdf	52999 83524c4c00f9f38df87befac1e94ef0d41e58 8be	no	24
Warnings:					
Information:					
5	Non Patent Literature	1999-IPCCC-Mineno_fonts.pdf	408040 f08887188e7e5dc3bef866d8ae624124e8cf 9161	no	7
Warnings:					
Information:					
6	Fee Worksheet (SB06)	fee-info.pdf	33654 f90cca2ff4817f8bc9e9b47539a3aeaa49c18f 546	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	1301372		

RED HAT 1010
RED HAT V. CAS

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

RED HAT 1010
RED HAT V. CAS

PTO/SB/06 (09-11)
Approved for use through 1/31/2014 OMB 0651-0032
U.S. Patent and Trademark Office, U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PATENT APPLICATION FEE DETERMINATION RECORD Substitute for Form PTO-875	Application or Docket Number 14/512,414	Filing Date 10/11/2014	☐ To be Mailed

ENTITY: ☐ LARGE ☒ SMALL ☐ MICRO

APPLICATION AS FILED – PART I

FOR	(Column 1) NUMBER FILED	(Column 2) NUMBER EXTRA	RATE ($)	FEE ($)
☐ BASIC FEE (37 CFR 1.16(a), (b), or (c))	N/A	N/A	N/A	
☐ SEARCH FEE (37 CFR 1.16(k), (l), or (m))	N/A	N/A	N/A	
☐ EXAMINATION FEE (37 CFR 1.16(o), (p), or (q))	N/A	N/A	N/A	
TOTAL CLAIMS (37 CFR 1.16(i))	minus 20 = *		X $ =	
INDEPENDENT CLAIMS (37 CFR 1.16(h))	minus 3 = *		X $ =	
☐ APPLICATION SIZE FEE (37 CFR 1.16(s))	If the specification and drawings exceed 100 sheets of paper, the application size fee due is $310 ($155 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).			
☐ MULTIPLE DEPENDENT CLAIM PRESENT (37 CFR 1.16(j))				
* If the difference in column 1 is less than zero, enter "0" in column 2.			TOTAL	

APPLICATION AS AMENDED – PART II

AMENDMENT **11/18/2015**

	(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	RATE ($)	ADDITIONAL FEE ($)
Total (37 CFR 1.16(i))	* 23	Minus	** 20	= 3	X $40 =	120
Independent (37 CFR 1.16(h))	* 5	Minus	***4	= 1	X $210 =	210
☐ Application Size Fee (37 CFR 1.16(s))						
☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))						
					TOTAL ADD'L FEE	330

AMENDMENT

	(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	RATE ($)	ADDITIONAL FEE ($)
Total (37 CFR 1.16(i))	*	Minus	**	=	X $ =	
Independent (37 CFR 1.16(h))	*	Minus	***	=	X $ =	
☐ Application Size Fee (37 CFR 1.16(s))						
☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))						
					TOTAL ADD'L FEE	

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.
** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20".
*** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3".
The "Highest Number Previously Paid For" (Total or Independent) is the highest number found in the appropriate box in column 1.

LIE
/VENICE WILLIAMS/

This collection of information is required by 37 CFR 1.16. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**
If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

RED HAT 1010
RED HAT V. CAS

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
14/512,414	10/11/2014	Eric M. Delangis	ED 2000	9530

64828 7590 12/30/2015

Harold McGurk
The Law Office of Clay McGurk
P.O. BOX 1488
Orange, CA 92856

EXAMINER	
CHO, HONG SOL	

ART UNIT	PAPER NUMBER
2467	

NOTIFICATION DATE	DELIVERY MODE
12/30/2015	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

CLAYMCGURK@GMAIL.COM

Office Action Summary	Application No. 14/512,414	Applicant(s) DELANGIS, ERIC M.	
	Examiner HONG CHO	Art Unit 2467	AIA (First Inventor to File) Status No

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTHS FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *11/18/2015*.

☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on _____.

2a)☒ This action is **FINAL**. 2b)☐ This action is non-final.

3)☐ An election was made by the applicant in response to a restriction requirement set forth during the interview on _____; the restriction requirement and election have been incorporated into this action.

4)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims*

5)☒ Claim(s) *1-23* is/are pending in the application.

5a) Of the above claim(s) _____ is/are withdrawn from consideration.

6)☐ Claim(s) _____ is/are allowed.

7)☒ Claim(s) *1-23* is/are rejected.

8)☐ Claim(s) _____ is/are objected to.

9)☐ Claim(s) _____ are subject to restriction and/or election requirement.

* If any claims have been determined allowable, you may be eligible to benefit from the **Patent Prosecution Highway** program at a participating intellectual property office for the corresponding application. For more information, please see http://www.uspto.gov/patents/init_events/pph/index.jsp or send an inquiry to PPHfeedback@uspto.gov.

Application Papers

10)☐ The specification is objected to by the Examiner.

11)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

Certified copies:

a)☐ All b)☐ Some** c)☐ None of the:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

** See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)

2) ☒ Information Disclosure Statement(s) (PTO/SB/08a and/or PTO/SB/08b) Paper No(s)/Mail Date _____.

3) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____.

4) ☐ Other: _____.

1. The present application is being examined under the pre-AIA first to invent provisions.

DETAILED ACTION

Response to Amendment

2. This office action is in response to the amendment filed on 11/18/2015. Claims 21-23 are added. Claims 1-23 are pending in the instant application.

Claim Rejections - 35 USC § 103

3. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

4. Claims 1-3, 6-8 and 11-23 are rejected under 35 U.S.C. 103(a) as being unpatentable over Perkins et al (US 6496477), hereinafter referred to as Perkins, in view of Li (US 20010037472) and Cain et al (US 8370507, "Cain").

RED HAT 1010
RED HAT V. CAS

Re claims 1, 7 and 21, Perkins receiving at the host (figure 12, element 1121) information about each of a plurality of devices (nodes in packet network 100) that are capable of joining a multipath connection with the initiating device (figure 12, element 1111) and the host performing at least one of (i) dividing data into groups of data packets and sending different ones of the groups of data packets to each one or more of the devices that are participating in the multipath connection, (ii) sending separate streams of data to each one or more of the devices that are participating in the multipath connection, and (iii) sending combinations of separate streams and groups of data packets to one or more of the devices that are participating in the multipath connection (column 24, lines 26-48). Perkins discloses multiple connections using path diversity via proxy servers (figure 22), but fails to disclose each of the network-edge devices being connected to the data-requesting device over a different connection and fails to disclose different data is sent over different connections from the network-edge devices to the data-requesting device. However, Li discloses using multiple links for providing multicast communication (figure 1, paragraph [0022]) and Cain discloses distributing multicast service to different multicast groups (column 1, lines 64-66). Therefore, it would have been obvious to one having ordinary skill in the art at the time the invention was made to modify the system of Perkins with Li and Cain for the benefit of providing reliable multicasting service.

Re claims 2, 3, 7, 8, 12 and 14, Perkins discloses the host associating a connection ID for each connection multipath request associated with each of the multipath connections and the host sending data associated with the particular connection ID to one

RED HAT 1010
RED HAT V. CAS

or more of the devices participating in the particular multipath connection associated with the particular connection ID (figure 12; multicasts communication stream based on group membership, column 24, lines 18-25).

Re claims 6, 11, 15, 16, 19 and 20, Perkins discloses the device is one of a multipath link aggregator, a network relay, and a multipath link server, a server, a network switch, a network router and a device capable of providing multipath services (figure 12, elements 1111, 1121 and 1125).

Re claims 13 and 18, Perkins discloses identifying one or more remote network devices to participate in capable of joining a multipath connection with the device, and receiving data from the host at the device via one or more of the identified network devices during the multipath connection (figure 12; column 24, lines 26-48).

Re claim 17, Perkins inherently discloses relaying data from another remote network device to the device through the selected one or more remote network devices (figure 12).

Re claims 22 and 23, Perkins discloses the first network-edge device is connected to a network, and wherein the second network-edge device is connected to the network or a different network, wherein the first connection and the second connection are wireless connections, a wired connection or a combination of wired and wireless connections (figure 22).

5. Claims 4, 5, 9 and 10 are rejected under 35 U.S.C. 103(a) as being unpatentable over Perkins in view of Li and Pung et al (US 20020150099), hereinafter referred to as Pung.

RED HAT 1010
RED HAT V. CAS

Re claims 4, 5, 9 and 10, Perkins discloses all of the limitations of the base claim, but fails to disclose receiving at the host information about which devices are to be removed from the multipath connection, the devices to be removed being associated with the connection ID. However, Pung discloses receiving a prune-branch packet from a joined computing device and removing the branch from the multicast tree (paragraph [0079]; figure 8a). Therefore, it would have been obvious to one having ordinary skill in the art at the time the invention was made to modify the system of Perkins with Pung for the benefit of providing efficient utilization of network resources by removing an entry in multicast tree table.

Response to Arguments

6. Applicant's arguments with respect to claims have been considered but are moot in view of the new ground(s) of rejection.

Conclusion

7. Applicant's amendment necessitated the new ground(s) of rejection presented in this Office action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a). A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is

RED HAT 1010
RED HAT V. CAS

not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the date of this final action.

8. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Hong Cho whose telephone number is 571-272-3087. The examiner can normally be reached on Mon-Fri during 7 am to 4 pm.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Hassan Phillips can be reached on 571-272-3940. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

/Hong Cho/

Primary Examiner, Art Unit 2467

RED HAT 1010
RED HAT V. CAS

Notice of References Cited	Application/Control No. 14/512,414	Applicant(s)/Patent Under Reexamination DELANGIS, ERIC M.	
	Examiner HONG CHO	Art Unit 2467	Page 1 of 1

U.S. PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	CPC Classification	US Classification
*	A	US-2001/0037472 A1	11-2001	Li, Yunzhou	H04L12/66	714/4.1
*	B	US-8,370,507 B1	02-2013	Cain; Bradley	H04N21/25816	709/225
	C	US-				
	D	US-				
	E	US-				
	F	US-				
	G	US-				
	H	US-				
	I	US-				
	J	US-				
	K	US-				
	L	US-				
	M	US-				

FOREIGN PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Country	Name	CPC Classification
	N					
	O					
	P					
	Q					
	R					
	S					
	T					

NON-PATENT DOCUMENTS

*		Include as applicable: Author, Title Date, Publisher, Edition or Volume, Pertinent Pages)
	U	
	V	
	W	
	X	

*A copy of this reference is not being furnished with this Office action. (See MPEP § 707.05(a).)
Dates in MM-YYYY format are publication dates. Classifications may be US or foreign

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	14512414	DELANGIS, ERIC M.
	Examiner	Art Unit
	HONG CHO	2467

CPC- SEARCHED		
Symbol	Date	Examiner

CPC COMBINATION SETS - SEARCHED		
Symbol	Date	Examiner

US CLASSIFICATION SEARCHED			
Class	Subclass	Date	Examiner

SEARCH NOTES		
Search Notes	Date	Examiner
H04L12/5692;H04L45/24;H04W88/16;H04W8/26;H04W28/06;H04W40/02 ;H04W80/00;H04W84/12;H04W84/18;H04W72/04 (EAST text search)	8/4/2015	hc
updated search	12/21/2015	hc

INTERFERENCE SEARCH			
US Class/ CPC Symbol	US Subclass / CPC Group	Date	Examiner
		8/4/2015	hc

Index of Claims	Application/Control No.	Applicant(s)/Patent Under Reexamination
	14512414	DELANGIS, ERIC M.
	Examiner	Art Unit
	HONG CHO	2467

✓	Rejected	-	Cancelled	N	Non-Elected	A	Appeal
=	Allowed	÷	Restricted	I	Interference	O	Objected

☐ Claims renumbered in the same order as presented by applicant ☐ CPA ☐ T.D. ☐ R.1.47

CLAIM		DATE								
Final	Original	08/11/2015	12/21/2015							
	1	✓	✓							
	2	✓	✓							
	3	✓	✓							
	4	✓	✓							
	5	✓	✓							
	6	✓	✓							
	7	✓	✓							
	8	✓	✓							
	9	✓	✓							
	10	✓	✓							
	11	✓	✓							
	12	✓	✓							
	13	✓	✓							
	14	✓	✓							
	15	✓	✓							
	16	✓	✓							
	17	✓	✓							
	18	✓	✓							
	19	✓	✓							
	20	✓	✓							
	21		✓							
	22		✓							
	23		✓							

Doc code: IDS
Doc description: Information Disclosure Statement (IDS) Filed

PTO/SB/08a (01-10)
Approved for use through 07/31/2012. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT
(Not for submission under 37 CFR 1.99)

Application Number	14512414
Filing Date	2014-10-11
First Named Inventor	Eric Delangis
Art Unit	2467
Examiner Name	Hong Cho
Attorney Docket Number	ED 2000

U.S.PATENTS Remove

Examiner Initial*	Cite No	Patent Number	Kind Code[1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1	7224964		2007-05-29	Souissi et al.	
	2	7917129		2011-03-29	Souisse et al.	
	3	8442504		2013-05-14	Souissi et al.	
	4	8521142		2013-08-27	Souissi et al.	

If you wish to add additional U.S. Patent citation information please click the Add button. Add

U.S.PATENT APPLICATION PUBLICATIONS Remove

Examiner Initial*	Cite No	Publication Number	Kind Code[1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1					

If you wish to add additional U.S. Published Application citation information please click the Add button. Add

FOREIGN PATENT DOCUMENTS Remove

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2] i	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	T[5]

EFS Web 2.1.17

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /H.S.C/

RED HAT 1010
RED HAT V. CAS

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	14512414
	Filing Date	2014-10-11
	First Named Inventor	Eric Delangis
	Art Unit	2467
	Examiner Name	Hong Cho
	Attorney Docket Number	ED 2000

	1							☐

If you wish to add additional Foreign Patent Document citation information please click the Add button Add

NON-PATENT LITERATURE DOCUMENTS Remove

Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc), date, pages(s), volume-issue number(s), publisher, city and/or country where published.	T[5]
	1	"Multiple Paths Protocol for a Cluster Type Network", Hiroshi Mineno et al., 7 pages (1999)	☐
	2	"The PPP Multilink Protocol (MP)", RFC 1990, 24 pages (1996)	☐

If you wish to add additional non-patent literature document citation information please click the Add button Add

EXAMINER SIGNATURE

Examiner Signature	/HONG S CHO/	Date Considered	12/21/2015

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

EFS Web 2.1.17

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /H.S.C/

RED HAT 1010
RED HAT V. CAS

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Not for submission under 37 CFR 1.99)

Application Number	14512414
Filing Date	2014-10-11
First Named Inventor	Eric Delangis
Art Unit	2467
Examiner Name	Hong Cho
Attorney Docket Number	ED 2000

CERTIFICATION STATEMENT

Please see 37 CFR 1.97 and 1.98 to make the appropriate selection(s):

☐ That each item of information contained in the information disclosure statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(1).

OR

☐ That no item of information contained in the information disclosure statement was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the person signing the certification after making reasonable inquiry, no item of information contained in the information disclosure statement was known to any individual designated in 37 CFR 1.56(c) more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(2).

☐ See attached certification statement.

☒ The fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☐ A certification statement is not submitted herewith.

SIGNATURE

A signature of the applicant or representative is required in accordance with CFR 1.33, 10.18. Please see CFR 1.4(d) for the form of the signature.

Signature	/H C McGurk IV/	Date (YYYY-MM-DD)	2015-11-18
Name/Print	H C McGurk	Registration Number	34964

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 1 hour to complete, including gathering, preparing and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /H.S.C/

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these record s.

2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.

3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.

4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).

5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.

6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).

7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.

8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.

9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /H.S.C/

RED HAT 1010
RED HAT V. CAS

EAST Search History

EAST Search History (Prior Art)

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
S299	27	S298 and join$3 with multicastS3 and multi near (path or link)	US-PGPUB; USPAT; USOCR	OR	OFF	2015/12/21 07:40
S298	44859	H04L12/5692;H04L45/24;H04W88/16;H04W8/26;H04W28/06;H04W40/02;H04W80/00;H04W84/12;H04W84/18;H04W72/04.cpc.	US-PGPUB; USPAT	OR	OFF	2015/12/21 07:40
S300	12	((Eric) near2 (Delangis)).INV.	US-PGPUB; USPAT; USOCR	OR	OFF	2015/12/21 07:45
S302	95	S301 and @rlad<"20021015"	US-PGPUB; USPAT	OR	OFF	2015/12/21 08:00
S301	620	(request$3 or join$3) with (two or multiple or different or several) near multicast$3	US-PGPUB; USPAT; USOCR	OR	OFF	2015/12/21 08:00
S305	4	S304 and @rlad<"20021015"	US-PGPUB; USPAT	OR	OFF	2015/12/21 08:02
S304	54	(request$3 or join$3) with (two or multiple or different or several) near multicast$3 near group	US-PGPUB; USPAT; USOCR	OR	OFF	2015/12/21 08:02
S306	1	09/661273	US-PGPUB; USPAT	OR	OFF	2015/12/21 08:03
S310	11	S309 and @rlad<"20021015"	US-PGPUB; USPAT	OR	OFF	2015/12/21 09:41
S309	160	(request$3 or join$3) with (two or multiple or different or several) near multicast$3 near group$1	US-PGPUB; USPAT; USOCR	OR	OFF	2015/12/21 09:41
S312	2	S311 and (requestS3 or join$3) with (two or multiple or different or several) near multicast$3 near group$1	US-PGPUB; USPAT; USOCR	OR	OFF	2015/12/21 09:45
L3	3	l2 and @rlad<"20021015"	US-PGPUB; USPAT	OR	OFF	2015/12/21 10:22
L2	51	(several or multiple or backup or second$3) near (connection or link) same multicast$3 near group$1	US-PGPUB; USPAT; USOCR	OR	OFF	2015/12/21 10:22

12/21/2015 11:36:01 AM
C:\Users\hcho\Documents\EAST\Workspaces\14512414.wsp

RED HAT 1010
RED HAT V. CAS

To: CLAYMCGURK@GMAIL.COM,,
From: PAIR_eOfficeAction@uspto.gov
Cc: PAIR_eOfficeAction@uspto.gov
Subject: Private PAIR Correspondence Notification for Customer Number 64828

Dec 30, 2015 05:40:58 AM

Dear PAIR Customer:

Harold McGurk
The Law Office of Clay McGurk
P.O. BOX 1488
Orange, CA 92856
UNITED STATES

The following USPTO patent application(s) associated with your Customer Number, 64828 , have new outgoing correspondence. This correspondence is now available for viewing in Private PAIR.

The official date of notification of the outgoing correspondence will be indicated on the form PTOL-90 accompanying the correspondence.

Disclaimer:
The list of documents shown below is provided as a courtesy and is not part of the official file wrapper. The content of the images shown in PAIR is the official record.

Application	Document	Mailroom Date	Attorney Docket No.
14512414	CTFR	12/30/2015	ED 2000
	892	12/30/2015	ED 2000
	1449	12/30/2015	ED 2000

To view your correspondence online or update your email addresses, please visit us anytime at https://sportal.uspto.gov/secure/myportal/privatepair.

If you have any questions, please email the Electronic Business Center (EBC) at EBC@uspto.gov with 'e-Office Action' on the subject line or call 1-866-217-9197 during the following hours:

Monday - Friday 6:00 a.m. to 12:00 a.m.

Thank you for prompt attention to this notice,

UNITED STATES PATENT AND TRADEMARK OFFICE

PATENT APPLICATION INFORMATION RETRIEVAL SYSTEM

RED HAT 1010
RED HAT V. CAS

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION NO.:	**14/512,414**	EXAMINER:	HONG CHO
INVENTOR:	Eric M. Delangis	GROUP ART UNIT:	2467
FILED:	OCTOBER 11, 2014	CONFIRMATION NO.:	9530
TITLE:	Multipath Communication Devices and Methods		
ATTORNEY NO:	ED 2000		

Commissioner For Patents
PO Box 1450
Alexandria, VA 22313-1450

RESPONSE UNDER 37 C.F.R. § 1.116

In response to the Office Action issued on December 30, 2015 ("Office Action"), please reconsider the above-identified application in light of the following remarks.

CERTIFICATE OF ELECTRONIC (EFS-WEB) TRANSMISSION
I hereby certify that this correspondence is being transmitted via the U.S. Patent Office's electronic filing system in accordance with 37 C.F.R. § 1.8(a)(1) from the **Pacific Time Zone** of the United States on the local date shown below.

Dated: March 30, 2016 By: Clay McGurk
Clay McGurk

MAILBOX AF

CLAIMS

This listing of claims will replace all prior versions and listings of the claims in the application:

1. (Previously Presented) A method for establishing and sending data during a multipath connection comprising:

a data-source device receiving multipath information about each of a plurality of network-edge devices that is capable of joining a multipath connection with a data-requesting device, each of the network-edge devices being connected to one or more networks, and each of the network-edge devices being connected to the data-requesting device over a different connection; and

the data-source device performing at least one of (i) sending different groups of data packets to different network-edge devices that are participating in the multipath connection, (ii) sending different streams of data to one or more of the network-edge devices that are participating in the multipath connection, and (iii) sending combinations of different streams and different groups of data packets to one or more of the network-edge devices that are participating in the multipath connection.

2. (Previously Presented) The method of claim 1, wherein the multipath information includes a connection ID that is generated by one of the data-source device, one of the plurality of network-edge devices, the data-requesting device and a network device.

3. (Previously Presented) The method of claim 1, further comprising:

the data-source device receiving second multipath information about a second network-edge device that is capable of joining the multipath connection; and

the data-source device joining the second network-edge device to the multipath connection.

RED HAT 1010
RED HAT V. CAS

4. (Previously Presented) The method of claim 1, further comprising:

the data-source device receiving multipath information about which network-edge devices are to be removed from the multipath connection; and

the data-source device removing the network-edge devices from being associated with the multipath connection.

5. (Previously Presented) The method of claim 1, further comprising at least one of the data-source device, the network-edge devices and the data-requesting device determining which of the network-edge devices are to be removed from the multipath connection.

6. (Previously Presented) The method of claim 1, wherein the data-source device is one of a multipath link aggregator, a network relay, a multipath link server, a server, a network switch, a network router and a device capable of providing multipath services.

7. (Previously Presented) A method for establishing and sending data during a multipath connection, comprising:

the data-source device receiving multipath connection requests from a plurality of data-requesting devices, each of the multipath connection requests involving a particular multipath connection to a data-requesting device;

the data-source device associating a connection ID to each multipath connection request;

the data-source device using multipath information about each of a plurality of network-edge devices that is capable of joining in a particular multipath connection to a particular data-requesting device, the particular multipath connection having a particular connection ID;

the data-source device notifying each of the network-edge devices that is participating in the particular multipath connection associated with the particular connection ID; and

the data-source device sending data associated with the particular connection ID to the network-edge devices participating in the particular multipath connection associated with the particular connection ID.

RED HAT 1010
RED HAT V. CAS

8. (Previously Presented) The method of claim 7, further comprising: the data-source device sending different groups of data packets to different network-edge devices that are participating in the particular multipath connection associated with the particular connection ID.

9. (Previously Presented) The method of claim 7, further comprising:

the data-source device receiving identification of the network-edge devices to be removed from the particular multipath connection; and

the data-source device removing the identified network-edge devices from the particular multipath connection.

10. (Previously Presented) The method of claim 7, further comprising the data-source device notifying the data-requesting device when all data has been sent over the particular multilink connection.

11. (Previously Presented) The method of claim 7, wherein the data-source device is one of a multipath link aggregator, a network relay, a multipath link server, a server, a network switch, a network router and a device capable of providing multipath services.

12. (Previously Presented) The method of claim 7, wherein the connection ID is generated by one of the data-source device, one of the plurality of network-edge devices, the data-requesting device, and a network device.

RED HAT 1010
RED HAT V. CAS

13. (Previously Presented) A method for establishing and sending data during a multipath connection, comprising:

a data-requesting device sending a multipath request to a network-edge device where the multipath request indicates that the network-edge device will join a multipath connection associated with the data-requesting device, the data-requesting device being connected over a first connection to a network and over a second connection to the network-edge device, the multipath connection comprising the first connection and the second connection which are capable of receiving data at the same time;

the data-requesting device receiving first data over the first connection from the network; and

after the network-edge device joins the multipath connection, the data-requesting device receiving second data over the second connection from the network-edge device.

14. (Previously Presented) The method of claim 13, wherein the first connection and the second connection are each a wireless connection, a wired connection or a combination of wired and wireless connections.

15. (Previously Presented) The method of claim 13, wherein the data-requesting device is one of a cell phone, a tablet, a switch, a network device, a router, a computer, a Wi-Fi device, a Wi-Max device, a communications device, a mobile communications device, a near-field communications device, and a Bluetooth device.

16. (Previously Presented) The method of claim 13, wherein the network-edge device comprises one or more routers, modems, cell phones, tablets, computers, cellular towers, wireless towers, base stations, wireless base stations, repeaters, wireless repeaters, Wi-Fi devices, networking devices, hot spot devices, Wi-MAX devices, LTE devices, GSM devices, consumer electronic devices, an Internet-ready device, a network switch, a network router, a multipath server and a device capable of providing multipath services.

RED HAT 1010
RED HAT V. CAS

17. (Previously Presented) The method of claim 13, wherein the data-requesting device is capable of connecting to another network device which is capable of joining the multipath connection, and further comprising the data-requesting device receiving data from the another network device after the another network device joins the multipath connection.

18. (Previously Presented) A data-requesting device comprising:

at least one interface capable of connecting concurrently to a plurality of network-edge devices during a multipath connection, the multipath connection comprising the connections between the data-requesting device and each of the network-edge devices; and

at least one processor coupled to the at least one interface that

sends multipath information to each of the network-edge devices that is capable of joining the multipath connection; and

receives different data at the data-requesting device from different ones of the network-edge devices that are able to join the multipath connection.

19. (Previously Presented) The data-requesting device of claim 18, wherein the data-requesting device is one of a cell phone, a tablet, a switch, a network device, a router, a computer, a Wi-Fi device, a Wi-Max device, a mobile communications device, a near-field communications device, and a Bluetooth device.

20. (Previously Presented) The data-requesting device of claim 18, wherein each of the network-edge devices comprises one or more network devices, communications devices, cellular devices, tablet devices, router devices, computer devices, satellite devices, Wi-MAX devices, wireless devices, mobile communications devices, Bluetooth devices, and Wi-Fi devices, a network switch, a network router, a multipath server and a device capable of providing multipath services.

RED HAT 1010
RED HAT V. CAS

21. (Previously Presented) A method for creating and using a multipath connection, comprising the steps of:

a data-requesting device sending first multipath control information to or through a first network-edge device, the first multipath control information indicates that the first network-edge device will join the multipath connection, the data-requesting device being connected over a first connection to the first network-edge device;

the data-requesting device sending a second multipath control information to or through a second network-edge device, the second multipath control information indicates that the second network-edge device will join the multipath connection, the data-requesting device being connected over a second connection to the second network-edge device, the data-requesting device capable of receiving different data over the first connection and the second connection at the same time; and

after the first network-edge device joins the multipath connection, the data-requesting device receiving first data over the first connection from the first network-edge device; and

after the second network-edge device joins the multipath connection, receiving second data at the data-requesting device over the second path from the second network-edge device.

22. (Previously Presented) The method as in claim 21, wherein the first network-edge device is connected to a network, and wherein the second network-edge device is connected to the network or a different network.

23. (Previously Presented) The method as in claim 21, wherein the first connection and the second connection are wireless connections, a wired connection or a combination of wired and wireless connections.

REMARKS

Status of the Claims

Claims 1-23 were pending in this application before this Response and remain pending pursuant to this Response.

35 U.S.C. § 103 Rejections

Starting on page 2 of the Office Action, claims 1-3, 6-8 and 11-23 were rejected under 35 U.S.C. §103(a) as being unpatentable over "Perkins" (U.S. Patent 6,496,477) in view of "Li" (U.S. Publ. Appl. 2001-0037472) and "Cain" (U.S. Patent 8,370,507).

On page 3, lines 11-14 of the Office Action, it is stated that Perkins "fails to disclose each of the network-edge devices being connected to the data-requesting device over a different connection and fails to disclose different data is sent over different connections from the network-edge devices to the data-requesting device." It is believed that Li was used to "disclose each of the network-edge devices being connected to the data-requesting device over a different connection", while Cain was used to disclose different data being sent over different connections from the network-edge devices to the data-requesting device. As will be discussed below, Li and Cain each does not describe or show what is required by the claims.

Li was used to disclose multiple connections where "figure 1, paragraph [0022]" were cited in support of the multiple connections. (Office Action, p. 3, lines 14-15). Figure 1 of Li shows, and paragraphs 21-22 of Li describe that the networking device 103 is coupled via a primary link 133 to Provider 1 ("104"), and is also coupled via backup link 144 to Provider 2 ("105"). A host interface connects networking device 103 to multicast hosts M1 (155) and M2 (165). A multicast host M1 and M2 are each an information consumer. (Li, ¶ 4 – definition of "multicast host").

The networking device 103 of Li is comparable to a "network-edge device" in the claims rather than the claims' "data-requesting device". In Li, it states that "the networking device receives a request from a multicast host to become a member of a multicast group and joins the multicast group over the primary interface on behalf of the multicast host." (Li, ¶ 10). It is the multicast hosts M1 (155) and M2 (165) that request the networking device 103 to join a particular multicast group. The networking device 103 is not making any requests to join the multicast group, rather these requests for multicast group data is instead being made by the

RED HAT 1010
RED HAT V. CAS

multicast hosts M1 and M2. Thus, the networking device 103 of Li is comparable to the "network-edge device" in the claims because it is not the device requesting data, such as made by the "data-requesting device" required by the claims.

There is nothing shown or described in Li that shows multiple networking devices 103 connecting to multicast hosts M1 (155) or M2 (165). Li is deficient like Perkins because both references only show and describe a single host interface between networking device 103 and multicast hosts M1 and M2. Li fails to disclose different "network-edge devices" being connected to the "data-requesting device" over different connections as required in the claims. Li would need to show multiple networking devices 103 connected over different interfaces to multicast hosts M1 and M2 - there is nothing that shows such a configuration in Li.

On page 3, lines 15-16 of the Office Action, Cain was used to disclose "distributing multicast service to different multicast groups" as cited in Cain 1:64-66: "[f]or internet television, television channels are sent to various subscribers over an IP network using IP multicast groups." Although Cain describes that different data is being sent to different multicast groups, Cain fails to make up for the deficiencies of Perkins and Li where multiple connections are required between a data-requesting device and a network-edge devices. Thus, all three references, whether individually or in combination, fail to describe the multiple connections between the data-requesting device and the network-edge devices as required in independent claims 1, 7, 13, 18 and 21. Any claim that is dependent on those allowable, independent claims is similarly allowable for the same reasons.

For these reasons, claims 1-3, 6-8 and 11-23 are patentably distinct and not anticipated by Perkins, Li and Cain either individually or in combination. It is respectfully requested that the rejection of claims 1-3, 6-8 and 11-23 under 35 U.S.C. §103 based on Perkins in view of Li and Cain be reconsidered and withdrawn.

35 U.S.C. § 103 Rejections

Starting on page 4 of the Office Action, claims 4, 5, 9 and 10 were rejected under 35 U.S.C. §102(e) as being unpatentable over Perkins, Li and Cain in view of "Pung" (U.S. Publication 2002-0150099). Pung suffers from the same deficiencies as Perkins, Li and Cain in that Fig. 1 of Pung shows that there is one only connection between the network and the data-requesting device. Claims 4, 5, 9 and 10 are allowable for the same reasons as provided above

RED HAT 1010
RED HAT V. CAS

for their corresponding amended independent claim. For this reason and those reasons distinguishing Perkins, Li and Cain, it is respectfully requested that the rejection of claims 4, 5, 9 and 10 under 35 U.S.C. § 103 based on Perkins, Li, Cain and Pung be reconsidered and withdrawn.

Conclusion

It is noted that claimed subject matter may be patentably distinguished from the cited references for additional reasons. It is also noted that any failure to comment directly upon any of the positions asserted in the Office Action does not indicate agreement or acquiescence with those asserted positions or allegations. Instead, it is believed that the foregoing sufficiently addresses the issues raised in the Office Action, and that the present claims are in condition for allowance.

All rejections in the Office Action have been properly traversed. It is respectfully requested that all outstanding rejections be reconsidered and withdrawn. It is believed that a full and complete reply has been made in response to the Office Action, and, that as such, the present patent application is in a condition for allowance.

The Examiner is always welcome to telephone the attorney of record at the telephone number provided below if such telephone discussion would help expedite the prosecution of the patent application. Prompt and favorable consideration of this Response is respectfully requested, making this application in a condition for allowance.

Respectfully Submitted,

Date: March 30, 2016

H C McGurk IV, Reg. No. 34,964
Attorney for Applicant

The Law Office of Clay McGurk
PO Box 1488
Orange, CA 92706
Phone: **949.395.3942**
Email: claymcgurk@gmail.com

Doc code: IDS
Doc description: Information Disclosure Statement (IDS) Filed

PTO/SB/08a (01-10)
Approved for use through 07/31/2012. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT
(Not for submission under 37 CFR 1.99)

Application Number	14512414
Filing Date	2014-10-11
First Named Inventor	Eric Delangis
Art Unit	2652
Examiner Name	
Attorney Docket Number	ED 2000

U.S.PATENTS Remove

Examiner Initial*	Cite No	Patent Number	Kind Code[1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1	5487069		1996-01-23	O'Sullivan et al.	

If you wish to add additional U.S. Patent citation information please click the Add button. Add

U.S.PATENT APPLICATION PUBLICATIONS Remove

Examiner Initial*	Cite No	Publication Number	Kind Code[1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1					

If you wish to add additional U.S. Published Application citation information please click the Add button. Add

FOREIGN PATENT DOCUMENTS Remove

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2]i	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	T[5]
	1							

If you wish to add additional Foreign Patent Document citation information please click the Add button Add

NON-PATENT LITERATURE DOCUMENTS Remove

Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc), date, pages(s), volume-issue number(s), publisher, city and/or country where published.	T[5]

EFS Web 2.1.17

RED HAT 1010
RED HAT V. CAS

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	14512414
	Filing Date	2014-10-11
	First Named Inventor	Eric Delangis
	Art Unit	2652
	Examiner Name	
	Attorney Docket Number	ED 2000

	1		

If you wish to add additional non-patent literature document citation information please click the Add button Add

EXAMINER SIGNATURE

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	14512414
	Filing Date	2014-10-11
	First Named Inventor	Eric Delangis
	Art Unit	2652
	Examiner Name	
	Attorney Docket Number	ED 2000

CERTIFICATION STATEMENT

Please see 37 CFR 1.97 and 1.98 to make the appropriate selection(s):

That each item of information contained in the information disclosure statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(1).

OR

☒ That no item of information contained in the information disclosure statement was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the person signing the certification after making reasonable inquiry, no item of information contained in the information disclosure statement was known to any individual designated in 37 CFR 1.56(c) more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(2).

See attached certification statement.

The fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

A certification statement is not submitted herewith.

SIGNATURE

A signature of the applicant or representative is required in accordance with CFR 1.33, 10.18. Please see CFR 1.4(d) for the form of the signature.

Signature	/H C McGurk IV/	Date (YYYY-MM-DD)	2016-03-30
Name/Print	H C McGurk	Registration Number	34964

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 1 hour to complete, including gathering, preparing and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these record s.

2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.

3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.

4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).

5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.

6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).

7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.

8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.

9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

Electronic Acknowledgement Receipt

EFS ID:	25351787
Application Number:	14512414
International Application Number:	
Confirmation Number:	9530
Title of Invention:	MULTIPATH COMMUNICATION DEVICES AND METHODS
First Named Inventor/Applicant Name:	Eric M. Delangis
Customer Number:	64828
Filer:	Harold C. McGurk
Filer Authorized By:	
Attorney Docket Number:	ED 2000
Receipt Date:	30-MAR-2016
Filing Date:	11-OCT-2014
Time Stamp:	18:08:56
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Response After Final Action	14512414_Reply_3-30-2016.pdf	122191 8618d46c19c0742bf823486e46bf4085611fc74	no	10
Warnings:					
Information:					

RED HAT 1010
RED HAT V. CAS

2	Information Disclosure Statement (IDS) Form (SB08)	14512414_IDS_3-30-2016.pdf	612145 60b9ba7993b9dd528ca87f60fc0207d1c72fa230	no	4
Warnings:					
Information:					
			Total Files Size (in bytes):	734336	

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111
If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371
If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office
If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

RED HAT 1010
RED HAT V. CAS

PTO/SB/06 (09-11)
Approved for use through 1/31/2014 OMB 0651-0032
U.S. Patent and Trademark Office, U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PATENT APPLICATION FEE DETERMINATION RECORD Substitute for Form PTO-875	Application or Docket Number 14/512,414	Filing Date 10/11/2014	☐ To be Mailed

ENTITY: ☐ LARGE ☒ SMALL ☐ MICRO

APPLICATION AS FILED – PART I

FOR	(Column 1) NUMBER FILED	(Column 2) NUMBER EXTRA	RATE ($)	FEE ($)
☐ BASIC FEE (37 CFR 1.16(a), (b), or (c))	N/A	N/A	N/A	
☐ SEARCH FEE (37 CFR 1.16(k), (l), or (m))	N/A	N/A	N/A	
☐ EXAMINATION FEE (37 CFR 1.16(o), (p), or (q))	N/A	N/A	N/A	
TOTAL CLAIMS (37 CFR 1.16(i))	minus 20 = *		X $ =	
INDEPENDENT CLAIMS (37 CFR 1.16(h))	minus 3 = *		X $ =	
☐ APPLICATION SIZE FEE (37 CFR 1.16(s))	If the specification and drawings exceed 100 sheets of paper, the application size fee due is $310 ($155 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).			
☐ MULTIPLE DEPENDENT CLAIM PRESENT (37 CFR 1.16(j))				
* If the difference in column 1 is less than zero, enter "0" in column 2.			TOTAL	

APPLICATION AS AMENDED – PART II

AMENDMENT	03/30/2016	(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	RATE ($)	ADDITIONAL FEE ($)
	Total (37 CFR 1.16(i))	* 23	Minus	** 23	= 0	X $40 =	0
	Independent (37 CFR 1.16(h))	* 5	Minus	***5	= 0	X $210 =	0
	☐ Application Size Fee (37 CFR 1.16(s))						
	☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))						
						TOTAL ADD'L FEE	0

AMENDMENT		(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	RATE ($)	ADDITIONAL FEE ($)
	Total (37 CFR 1.16(i))	*	Minus	**	=	X $ =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=	X $ =	
	☐ Application Size Fee (37 CFR 1.16(s))						
	☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))						
						TOTAL ADD'L FEE	

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.
** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20".
*** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3".
The "Highest Number Previously Paid For" (Total or Independent) is the highest number found in the appropriate box in column 1.

LIE
/JACQUELYN WILLIAMS/

This collection of information is required by 37 CFR 1.16. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**
If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

RED HAT 1010
RED HAT V. CAS

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

NOTICE OF ALLOWANCE AND FEE(S) DUE

64828 7590 04/13/2016

Harold McGurk
The Law Office of Clay McGurk
P.O. BOX 1488
Orange, CA 92856

EXAMINER
CHO, HONG SOL

ART UNIT	PAPER NUMBER
2467	

DATE MAILED: 04/13/2016

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
14/512,414	10/11/2014	Eric M. Delangis	ED 2000	9530

TITLE OF INVENTION: MULTIPATH COMMUNICATION DEVICES AND METHODS

APPLN. TYPE	ENTITY STATUS	ISSUE FEE DUE	PUBLICATION FEE DUE	PREV. PAID ISSUE FEE	TOTAL FEE(S) DUE	DATE DUE
nonprovisional	SMALL	$480	$0	$0	$480	07/13/2016

THE APPLICATION IDENTIFIED ABOVE HAS BEEN EXAMINED AND IS ALLOWED FOR ISSUANCE AS A PATENT. PROSECUTION ON THE MERITS IS CLOSED. THIS NOTICE OF ALLOWANCE IS NOT A GRANT OF PATENT RIGHTS. THIS APPLICATION IS SUBJECT TO WITHDRAWAL FROM ISSUE AT THE INITIATIVE OF THE OFFICE OR UPON PETITION BY THE APPLICANT. SEE 37 CFR 1.313 AND MPEP 1308.

THE ISSUE FEE AND PUBLICATION FEE (IF REQUIRED) MUST BE PAID WITHIN THREE MONTHS FROM THE MAILING DATE OF THIS NOTICE OR THIS APPLICATION SHALL BE REGARDED AS ABANDONED. THIS STATUTORY PERIOD CANNOT BE EXTENDED. SEE 35 U.S.C. 151. THE ISSUE FEE DUE INDICATED ABOVE DOES NOT REFLECT A CREDIT FOR ANY PREVIOUSLY PAID ISSUE FEE IN THIS APPLICATION. IF AN ISSUE FEE HAS PREVIOUSLY BEEN PAID IN THIS APPLICATION (AS SHOWN ABOVE), THE RETURN OF PART B OF THIS FORM WILL BE CONSIDERED A REQUEST TO REAPPLY THE PREVIOUSLY PAID ISSUE FEE TOWARD THE ISSUE FEE NOW DUE.

HOW TO REPLY TO THIS NOTICE:

I. Review the ENTITY STATUS shown above. If the ENTITY STATUS is shown as SMALL or MICRO, verify whether entitlement to that entity status still applies.

If the ENTITY STATUS is the same as shown above, pay the TOTAL FEE(S) DUE shown above.

If the ENTITY STATUS is changed from that shown above, on PART B - FEE(S) TRANSMITTAL, complete section number 5 titled "Change in Entity Status (from status indicated above)".

For purposes of this notice, small entity fees are 1/2 the amount of undiscounted fees, and micro entity fees are 1/2 the amount of small entity fees.

II. PART B - FEE(S) TRANSMITTAL, or its equivalent, must be completed and returned to the United States Patent and Trademark Office (USPTO) with your ISSUE FEE and PUBLICATION FEE (if required). If you are charging the fee(s) to your deposit account, section "4b" of Part B - Fee(s) Transmittal should be completed and an extra copy of the form should be submitted. If an equivalent of Part B is filed, a request to reapply a previously paid issue fee must be clearly made, and delays in processing may occur due to the difficulty in recognizing the paper as an equivalent of Part B.

III. All communications regarding this application must give the application number. Please direct all communications prior to issuance to Mail Stop ISSUE FEE unless advised to the contrary.

IMPORTANT REMINDER: Utility patents issuing on applications filed on or after Dec. 12, 1980 may require payment of maintenance fees. It is patentee's responsibility to ensure timely payment of maintenance fees when due.

PART B - FEE(S) TRANSMITTAL

Complete and send this form, together with applicable fee(s), to: Mail **Mail Stop ISSUE FEE**
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450
or Fax (571)-273-2885

INSTRUCTIONS: This form should be used for transmitting the ISSUE FEE and PUBLICATION FEE (if required). Blocks 1 through 5 should be completed where appropriate. All further correspondence including the Patent, advance orders and notification of maintenance fees will be mailed to the current correspondence address as indicated unless corrected below or directed otherwise in Block 1, by (a) specifying a new correspondence address; and/or (b) indicating a separate "FEE ADDRESS" for maintenance fee notifications.

CURRENT CORRESPONDENCE ADDRESS (Note: Use Block 1 for any change of address)

64828 7590 04/13/2016
Harold McGurk
The Law Office of Clay McGurk
P.O. BOX 1488
Orange, CA 92856

Note: A certificate of mailing can only be used for domestic mailings of the Fee(s) Transmittal. This certificate cannot be used for any other accompanying papers. Each additional paper, such as an assignment or formal drawing, must have its own certificate of mailing or transmission.

Certificate of Mailing or Transmission
I hereby certify that this Fee(s) Transmittal is being deposited with the United States Postal Service with sufficient postage for first class mail in an envelope addressed to the Mail Stop ISSUE FEE address above, or being facsimile transmitted to the USPTO (571) 273-2885, on the date indicated below.

(Depositor's name)
(Signature)
(Date)

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
14/512,414	10/11/2014	Eric M. Delangis	ED 2000	9530

TITLE OF INVENTION: MULTIPATH COMMUNICATION DEVICES AND METHODS

APPLN. TYPE	ENTITY STATUS	ISSUE FEE DUE	PUBLICATION FEE DUE	PREV. PAID ISSUE FEE	TOTAL FEE(S) DUE	DATE DUE
nonprovisional	SMALL	$480	$0	$0	$480	07/13/2016

EXAMINER	ART UNIT	CLASS-SUBCLASS
CHO, HONG SOL	2467	370-235000

1. Change of correspondence address or indication of "Fee Address" (37 CFR 1.363).
- ☐ Change of correspondence address (or Change of Correspondence Address form PTO/SB/122) attached.
- ☐ "Fee Address" indication (or "Fee Address" Indication form PTO/SB/47; Rev 03-02 or more recent) attached. **Use of a Customer Number is required.**

2. For printing on the patent front page, list
(1) The names of up to 3 registered patent attorneys or agents OR, alternatively,
(2) The name of a single firm (having as a member a registered attorney or agent) and the names of up to 2 registered patent attorneys or agents. If no name is listed, no name will be printed.
1 _____
2 _____
3 _____

3. ASSIGNEE NAME AND RESIDENCE DATA TO BE PRINTED ON THE PATENT (print or type)

PLEASE NOTE: Unless an assignee is identified below, no assignee data will appear on the patent. If an assignee is identified below, the document has been filed for recordation as set forth in 37 CFR 3.11. Completion of this form is NOT a substitute for filing an assignment.

(A) NAME OF ASSIGNEE (B) RESIDENCE: (CITY and STATE OR COUNTRY)

Please check the appropriate assignee category or categories (will not be printed on the patent) : ☐ Individual ☐ Corporation or other private group entity ☐ Government

4a. The following fee(s) are submitted:
- ☐ Issue Fee
- ☐ Publication Fee (No small entity discount permitted)
- ☐ Advance Order - # of Copies

4b. Payment of Fee(s): **(Please first reapply any previously paid issue fee shown above)**
- ☐ A check is enclosed.
- ☐ Payment by credit card. Form PTO-2038 is attached.
- ☐ The director is hereby authorized to charge the required fee(s), any deficiency, or credits any overpayment, to Deposit Account Number _____ (enclose an extra copy of this form).

5. **Change in Entity Status** (from status indicated above)
- ☐ Applicant certifying micro entity status. See 37 CFR 1.29
  NOTE: Absent a valid certification of Micro Entity Status (see forms PTO/SB/15A and 15B), issue fee payment in the micro entity amount will not be accepted at the risk of application abandonment.
- ☐ Applicant asserting small entity status. See 37 CFR 1.27
  NOTE: If the application was previously under micro entity status, checking this box will be taken to be a notification of loss of entitlement to micro entity status.
- ☐ Applicant changing to regular undiscounted fee status.
  NOTE: Checking this box will be taken to be a notification of loss of entitlement to small or micro entity status, as applicable.

NOTE: This form must be signed in accordance with 37 CFR 1.31 and 1.33. See 37 CFR 1.4 for signature requirements and certifications.

Authorized Signature _____ Date _____

Typed or printed name Registration No.

PTOL-85 Part B (10-13) Approved for use through 10/31/2013. OMB 0651-0033 U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

RED HAT 1010
RED HAT V. CAS

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
14/512,414	10/11/2014	Eric M. Delangis	ED 2000	9530

64828 7590 04/13/2016

Harold McGurk
The Law Office of Clay McGurk
P.O. BOX 1488
Orange, CA 92856

EXAMINER	
CHO, HONG SOL	

ART UNIT	PAPER NUMBER
2467	

DATE MAILED: 04/13/2016

Determination of Patent Term Adjustment under 35 U.S.C. 154 (b)
(Applications filed on or after May 29, 2000)

The Office has discontinued providing a Patent Term Adjustment (PTA) calculation with the Notice of Allowance.

Section 1(h)(2) of the AIA Technical Corrections Act amended 35 U.S.C. 154(b)(3)(B)(i) to eliminate the requirement that the Office provide a patent term adjustment determination with the notice of allowance. See Revisions to Patent Term Adjustment, 78 Fed. Reg. 19416, 19417 (Apr. 1, 2013). Therefore, the Office is no longer providing an initial patent term adjustment determination with the notice of allowance. The Office will continue to provide a patent term adjustment determination with the Issue Notification Letter that is mailed to applicant approximately three weeks prior to the issue date of the patent, and will include the patent term adjustment on the patent. Any request for reconsideration of the patent term adjustment determination (or reinstatement of patent term adjustment) should follow the process outlined in 37 CFR 1.705.

Any questions regarding the Patent Term Extension or Adjustment determination should be directed to the Office of Patent Legal Administration at (571)-272-7702. Questions relating to issue and publication fee payments should be directed to the Customer Service Center of the Office of Patent Publication at 1-(888)-786-0101 or (571)-272-4200.

OMB Clearance and PRA Burden Statement for PTOL-85 Part B

The Paperwork Reduction Act (PRA) of 1995 requires Federal agencies to obtain Office of Management and Budget approval before requesting most types of information from the public. When OMB approves an agency request to collect information from the public, OMB (i) provides a valid OMB Control Number and expiration date for the agency to display on the instrument that will be used to collect the information and (ii) requires the agency to inform the public about the OMB Control Number's legal significance in accordance with 5 CFR 1320.5(b).

The information collected by PTOL-85 Part B is required by 37 CFR 1.311. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, Virginia 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, Virginia 22313-1450. Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether disclosure of these records is required by the Freedom of Information Act.
2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.
3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.
4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).
5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.
6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).
7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.
8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspection or an issued patent.
9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

RED HAT 1010
RED HAT V. CAS

Notice of Allowability	Application No. 14/512,414	Applicant(s) DELANGIS, ERIC M.	
	Examiner HONG CHO	Art Unit 2467	AIA (First Inventor to File) Status No

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address--

All claims being allowable, PROSECUTION ON THE MERITS IS (OR REMAINS) CLOSED in this application. If not included herewith (or previously mailed), a Notice of Allowance (PTOL-85) or other appropriate communication will be mailed in due course. **THIS NOTICE OF ALLOWABILITY IS NOT A GRANT OF PATENT RIGHTS.** This application is subject to withdrawal from issue at the initiative of the Office or upon petition by the applicant. See 37 CFR 1.313 and MPEP 1308.

1. ☒ This communication is responsive to *03/30/2016*.
   ☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on_____.
2. ☐ An election was made by the applicant in response to a restriction requirement set forth during the interview on _____; the restriction requirement and election have been incorporated into this action.
3. ☒ The allowed claim(s) is/are *1-23*. As a result of the allowed claim(s), you may be eligible to benefit from the **Patent Prosecution Highway** program at a participating intellectual property office for the corresponding application. For more information, please see http://www.uspto.gov/patents/init_events/pph/index.jsp or send an inquiry to PPHfeedback@uspto.gov .
4. ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

   Certified copies:

   a) ☐ All b) ☐ Some *c) ☐ None of the:
   1. ☐ Certified copies of the priority documents have been received.
   2. ☐ Certified copies of the priority documents have been received in Application No. _____ .
   3. ☐ Copies of the certified copies of the priority documents have been received in this national stage application from the International Bureau (PCT Rule 17.2(a)).

   * Certified copies not received: _____.

Applicant has **THREE MONTHS FROM THE** "MAILING DATE" of this communication to file a reply complying with the requirements noted below. Failure to timely comply will result in ABANDONMENT of this application.
THIS THREE-MONTH PERIOD IS NOT EXTENDABLE.

5. ☐ CORRECTED DRAWINGS (as "replacement sheets") must be submitted.
   ☐ including changes required by the attached Examiner's Amendment / Comment or in the Office action of Paper No./Mail Date _____.

   Identifying indicia such as the application number (see 37 CFR 1.84(c)) should be written on the drawings in the front (not the back) of each sheet. Replacement sheet(s) should be labeled as such in the header according to 37 CFR 1.121(d).
6. ☐ DEPOSIT OF and/or INFORMATION about the deposit of BIOLOGICAL MATERIAL must be submitted. Note the attached Examiner's comment regarding REQUIREMENT FOR THE DEPOSIT OF BIOLOGICAL MATERIAL.

Attachment(s)

1. ☐ Notice of References Cited (PTO-892)
2. ☒ Information Disclosure Statements (PTO/SB/08), Paper No./Mail Date _____
3. ☐ Examiner's Comment Regarding Requirement for Deposit of Biological Material
4. ☐ Interview Summary (PTO-413), Paper No./Mail Date _____ .
5. ☐ Examiner's Amendment/Comment
6. ☐ Examiner's Statement of Reasons for Allowance
7. ☐ Other _____.

/HONG CHO/
Primary Examiner, Art Unit 2467

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	14512414	DELANGIS, ERIC M.
	Examiner	Art Unit
	HONG CHO	2467

CPC- SEARCHED		
Symbol	Date	Examiner

CPC COMBINATION SETS - SEARCHED		
Symbol	Date	Examiner

US CLASSIFICATION SEARCHED			
Class	Subclass	Date	Examiner

SEARCH NOTES		
Search Notes	Date	Examiner
H04L12/5692;H04L45/24;H04W88/16;H04W8/26;H04W28/06;H04W40/02 ;H04W80/00;H04W84/12;H04W84/18;H04W72/04 (EAST text search)	8/4/2015	hc
updated search	12/21/2015	hc
updated search	4/6/2016	hc

INTERFERENCE SEARCH			
US Class/ CPC Symbol	US Subclass / CPC Group	Date	Examiner
	see search history printout	4/6/2016	hc

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

BIB DATA SHEET

CONFIRMATION NO. 9530

SERIAL NUMBER	FILING or 371(c) DATE RULE	CLASS	GROUP ART UNIT	ATTORNEY DOCKET NO.
14/512,414	10/11/2014	370	2467	ED 2000

APPLICANTS

INVENTORS

Eric M. Delangis, Huntsville, AL;

**** CONTINUING DATA ***************************

This application is a CON of 13/531,294 06/22/2012 PAT 8861349 *
which is a CON of 12/581,852 10/19/2009 PAT 8228801
which is a CON of 10/686,375 10/14/2003 PAT 7606156
which claims benefit of 60/418,521 10/15/2002
(*)Data provided by applicant is not consistent with PTO records.

**** FOREIGN APPLICATIONS ***************************

**** IF REQUIRED, FOREIGN FILING LICENSE GRANTED **** ** SMALL ENTITY **
10/16/2014

Foreign Priority claimed	☐ Yes ☑ No	☑ Met after Allowance	STATE OR COUNTRY	SHEETS DRAWINGS	TOTAL CLAIMS	INDEPENDENT CLAIMS
35 USC 119(a-d) conditions met	☐ Yes ☑ No					
Verified and Acknowledged	/HONG SOL CHO/ Examiner's Signature	hc Initials	AL	9	20	4

ADDRESS

Harold McGurk
The Law Office of Clay McGurk
P.O. BOX 1488
Orange, CA 92856
UNITED STATES

TITLE

MULTIPATH COMMUNICATION DEVICES AND METHODS

FILING FEE RECEIVED		
1270	FEES: Authority has been given in Paper No._____ to charge/credit DEPOSIT ACCOUNT No._____ for following:	☐ All Fees ☐ 1.16 Fees (Filing) ☐ 1.17 Fees (Processing Ext. of time) ☐ 1.18 Fees (Issue) ☐ Other _____ ☐ Credit

BIB (Rev. 05/07).

RED HAT 1010
RED HAT V. CAS

EAST Search History

EAST Search History (Prior Art)

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
L1	47451	H04L12/5692;H04L45/24;H04W88/16;H04W8/26;H04W28/06;H04W40/02;H04W80/00;H04W84/12;H04W84/18;H04W72/04.cpc.	US-PGPUB; USPAT	OR	OFF	2016/04/06 15:38
L2	31	L1 and join$3 with multicast$3 and multi near (path or link)	US-PGPUB; USPAT; USOCR	OR	OFF	2016/04/06 15:38
L3	1	"5487069".pn.	US-PGPUB; USPAT; USOCR	OR	OFF	2016/04/06 15:39
L8	12	((Eric) near2 (Delangis)).INV.	US-PGPUB; USPAT; USOCR	OR	OFF	2016/04/06 15:43
L9	54	(several or multiple or backup or second$3) near (connection or link) same multicast$3 near group$1	US-PGPUB; USPAT; USOCR	OR	OFF	2016/04/06 15:43
L10	56	(request$3 or join$3) with (two or multiple or different or several) near multicast$3 near group	US-PGPUB; USPAT; USOCR	OR	OFF	2016/04/06 16:01
L11	12	((requestS3 or joinS3) with (two or multiple or different or several) near multicast$3 near group).clm.	US-PGPUB; USPAT; USOCR	OR	OFF	2016/04/06 16:08

4/6/2016 4:21:28 PM
C:\Users\hcho\Documents\EAST\Workspaces\14512414.wsp

Doc code: IDS
Doc description: Information Disclosure Statement (IDS) Filed

PTO/SB/08a (01-10)
Approved for use through 07/31/2012. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Not for submission under 37 CFR 1.99)

Application Number	14512414
Filing Date	2014-10-11
First Named Inventor	Eric Delangis
Art Unit	2652
Examiner Name	
Attorney Docket Number	ED 2000

U.S.PATENTS Remove

Examiner Initial*	Cite No	Patent Number	Kind Code[1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1	5487069		1996-01-23	O'Sullivan et al.	

If you wish to add additional U.S. Patent citation information please click the Add button. Add

U.S.PATENT APPLICATION PUBLICATIONS Remove

Examiner Initial*	Cite No	Publication Number	Kind Code[1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1					

If you wish to add additional U.S. Published Application citation information please click the Add button. Add

FOREIGN PATENT DOCUMENTS Remove

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2] i	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	T[5]
	1							

If you wish to add additional Foreign Patent Document citation information please click the Add button Add

NON-PATENT LITERATURE DOCUMENTS Remove

Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc), date, pages(s), volume-issue number(s), publisher, city and/or country where published.	T[5]

EFS Web 2.1.17

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /H.S.C/

RED HAT 1010
RED HAT V. CAS

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	14512414
	Filing Date	2014-10-11
	First Named Inventor	Eric Delangis
	Art Unit	2652
	Examiner Name	
	Attorney Docket Number	ED 2000

	1		

If you wish to add additional non-patent literature document citation information please click the Add button [Add]

EXAMINER SIGNATURE

Examiner Signature	/HONG S CHO/	Date Considered	04/06/2016

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

EFS Web 2.1.17

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /H.S.C/

RED HAT 1010
RED HAT V. CAS

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Not for submission under 37 CFR 1.99)

Application Number	14512414
Filing Date	2014-10-11
First Named Inventor	Eric Delangis
Art Unit	2652
Examiner Name	
Attorney Docket Number	ED 2000

CERTIFICATION STATEMENT

Please see 37 CFR 1.97 and 1.98 to make the appropriate selection(s):

That each item of information contained in the information disclosure statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(1).

OR

☒ That no item of information contained in the information disclosure statement was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the person signing the certification after making reasonable inquiry, no item of information contained in the information disclosure statement was known to any individual designated in 37 CFR 1.56(c) more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(2).

See attached certification statement.

The fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

A certification statement is not submitted herewith.

SIGNATURE

A signature of the applicant or representative is required in accordance with CFR 1.33, 10.18. Please see CFR 1.4(d) for the form of the signature.

Signature	/H C McGurk IV/	Date (YYYY-MM-DD)	2016-03-30
Name/Print	H C McGurk	Registration Number	34964

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 1 hour to complete, including gathering, preparing and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /H.S.C/

RED HAT 1010
RED HAT V. CAS

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these record s.

2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.

3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.

4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).

5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.

6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).

7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.

8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.

9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

 ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /H.S.C/

RED HAT 1010
RED HAT V. CAS

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION NO.:	**14/512,414**	EXAMINER:	HONG CHO
INVENTOR:	Eric M. Delangis	GROUP ART UNIT:	2467
FILED:	OCTOBER 11, 2014	CONFIRMATION NO.:	9530
TITLE:	Multipath Communication Devices and Methods		
ATTORNEY NO:	ED 2000		

Commissioner For Patents
PO Box 1450
Alexandria, VA 22313-1450

OK TO ENTER: /H.S.C/

RESPONSE UNDER 37 C.F.R. § 1.116

In response to the Office Action issued on December 30, 2015 ("Office Action"), please reconsider the above-identified application in light of the following remarks.

CERTIFICATE OF ELECTRONIC (EFS-WEB) TRANSMISSION
I hereby certify that this correspondence is being transmitted via the U.S. Patent Office's electronic filing system in accordance with 37 C.F.R. § 1.8(a)(1) from the **Pacific Time Zone** of the United States on the local date shown below.

Dated: March 30, 2016

By: Clay McGurk
Clay McGurk

MAILBOX AF

Issue Classification	Application/Control No.	Applicant(s)/Patent Under Reexamination
	14512414	DELANGIS, ERIC M.
	Examiner	**Art Unit**
	HONG CHO	2467

CPC						
Symbol					**Type**	**Version**
H04L		45	/	24	F	2013-01-01
H04L		12	/	5692	I	2013-01-01
H04W		8	/	26	A	2013-01-01
H04W		28	/	06	A	2013-01-01
H04W		40	/	02	A	2013-01-01
H04W		72	/	04	A	2013-01-01
H04W		80	/	00	A	2013-01-01
H04W		84	/	12	A	2013-01-01
H04W		84	/	18	A	2013-01-01
H04W		88	/	16	I	2013-01-01
			/			
			/			
			/			
			/			
			/			

CPC Combination Sets							
Symbol					**Type**	**Set**	**Ranking**	**Version**
			/					
			/					

NONE		Total Claims Allowed:	
(Assistant Examiner)	(Date)	23	
/HONG CHO/ Primary Examiner.Art Unit 2467	04/06/2016	O.G. Print Claim(s)	O.G. Print Figure
(Primary Examiner)	(Date)	1	1

Issue Classification	Application/Control No.	Applicant(s)/Patent Under Reexamination
	14512414	DELANGIS, ERIC M.
	Examiner	Art Unit
	HONG CHO	2467

US ORIGINAL CLASSIFICATION		INTERNATIONAL CLASSIFICATION										
CLASS	SUBCLASS	CLAIMED					NON-CLAIMED					
		H	0	4	L	12 / 207						
CROSS REFERENCE(S)		H	0	4	L	12 / 54 (2013.01.01)						
		H	0	4	W	88 / 16 (2009.01.01)						
CLASS	SUBCLASS (ONE SUBCLASS PER BLOCK)	H	0	4	W	28 / 06 (2009.01.01)						

NONE (Assistant Examiner)	(Date)	Total Claims Allowed: 23	
/HONG CHO/ Primary Examiner.Art Unit 2467 (Primary Examiner)	04/06/2016 (Date)	O.G. Print Claim(s) 1	O.G. Print Figure 1

U.S. Patent and Trademark Office

Part of Paper No. 20160406

RED HAT 1010
RED HAT V. CAS

Issue Classification	Application/Control No.	Applicant(s)/Patent Under Reexamination
	14512414	DELANGIS, ERIC M.
	Examiner	**Art Unit**
	HONG CHO	2467

☒ Claims renumbered in the same order as presented by applicant ☐ CPA ☐ T.D. ☐ R.1.47

Final	Original	Final	Original	Final	Original	Final	Original	Final	Original	Final	Original	Final	Original	Final	Original

NONE (Assistant Examiner)	(Date)	**Total Claims Allowed:** 23	
/HONG CHO/ Primary Examiner.Art Unit 2467 (Primary Examiner)	04/06/2016 (Date)	O.G. Print Claim(s) 1	O.G. Print Figure 1

U.S. Patent and Trademark Office Part of Paper No. 20160406

RED HAT 1010
RED HAT V. CAS

To: CLAYMCGURK@GMAIL.COM,,
From: PAIR_eOfficeAction@uspto.gov
Cc: PAIR_eOfficeAction@uspto.gov
Subject: Private PAIR Correspondence Notification for Customer Number 64828

Apr 13, 2016 05:38:32 AM

Dear PAIR Customer:

Harold McGurk
The Law Office of Clay McGurk
P.O. BOX 1488
Orange, CA 92856
UNITED STATES

The following USPTO patent application(s) associated with your Customer Number, 64828 , have new outgoing correspondence. This correspondence is now available for viewing in Private PAIR.

The official date of notification of the outgoing correspondence will be indicated on the form PTOL-90 accompanying the correspondence.

Disclaimer:
The list of documents shown below is provided as a courtesy and is not part of the official file wrapper. The content of the images shown in PAIR is the official record.

Application	Document	Mailroom Date	Attorney Docket No.
14512414	NOA	04/13/2016	ED 2000
	1449	04/13/2016	ED 2000

To view your correspondence online or update your email addresses, please visit us anytime at https://sportal.uspto.gov/secure/myportal/privatepair.

If you have any questions, please email the Electronic Business Center (EBC) at EBC@uspto.gov with 'e-Office Action' on the subject line or call 1-866-217-9197 during the following hours:

Monday - Friday 6:00 a.m. to 12:00 a.m.

Thank you for prompt attention to this notice,

UNITED STATES PATENT AND TRADEMARK OFFICE

PATENT APPLICATION INFORMATION RETRIEVAL SYSTEM

RED HAT 1010
RED HAT V. CAS

PART B - FEE(S) TRANSMITTAL

Complete and send this form, together with applicable fee(s), to: Mail **Mail Stop ISSUE FEE**
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450
or Fax (571)-273-2885

INSTRUCTIONS: This form should be used for transmitting the ISSUE FEE and PUBLICATION FEE (if required). Blocks 1 through 5 should be completed where appropriate. All further correspondence including the Patent, advance orders and notification of maintenance fees will be mailed to the current correspondence address as indicated unless corrected below or directed otherwise in Block 1, by (a) specifying a new correspondence address; and/or (b) indicating a separate "FEE ADDRESS" for maintenance fee notifications.

CURRENT CORRESPONDENCE ADDRESS (Note: Use Block 1 for any change of address)

64828 7590 04/13/2016
Harold McGurk
The Law Office of Clay McGurk
P.O. BOX 1488
Orange, CA 92856

Note: A certificate of mailing can only be used for domestic mailings of the Fee(s) Transmittal. This certificate cannot be used for any other accompanying papers. Each additional paper, such as an assignment or formal drawing, must have its own certificate of mailing or transmission.

Certificate of Mailing or Transmission
I hereby certify that this Fee(s) Transmittal is being deposited with the United States Postal Service with sufficient postage for first class mail in an envelope addressed to the Mail Stop ISSUE FEE address above, or being facsimile transmitted to the USPTO (571) 273-2885, on the date indicated below.

(Depositor's name)
(Signature)
(Date)

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
14/512,414	10/11/2014	Eric M. Delangis	ED 2000	9530

TITLE OF INVENTION: MULTIPATH COMMUNICATION DEVICES AND METHODS

APPLN. TYPE	ENTITY STATUS	ISSUE FEE DUE	PUBLICATION FEE DUE	PREV. PAID ISSUE FEE	TOTAL FEE(S) DUE	DATE DUE
nonprovisional	SMALL	$480	$0	$0	$480	07/13/2016

EXAMINER	ART UNIT	CLASS-SUBCLASS
CHO, HONG SOL	2467	370-235000

1. Change of correspondence address or indication of "Fee Address" (37 CFR 1.363).
- ☐ Change of correspondence address (or Change of Correspondence Address form PTO/SB/122) attached.
- ☐ "Fee Address" indication (or "Fee Address" Indication form PTO/SB/47; Rev 03-02 or more recent) attached. **Use of a Customer Number is required.**

2. For printing on the patent front page, list
(1) The names of up to 3 registered patent attorneys or agents OR, alternatively,
(2) The name of a single firm (having as a member a registered attorney or agent) and the names of up to 2 registered patent attorneys or agents. If no name is listed, no name will be printed.

1 Clay McGurk
2 The Law Office of Clay McGurk
3

3. ASSIGNEE NAME AND RESIDENCE DATA TO BE PRINTED ON THE PATENT (print or type)

PLEASE NOTE: Unless an assignee is identified below, no assignee data will appear on the patent. If an assignee is identified below, the document has been filed for recordation as set forth in 37 CFR 3.11. Completion of this form is NOT a substitute for filing an assignment.

(A) NAME OF ASSIGNEE (B) RESIDENCE: (CITY and STATE OR COUNTRY)

Please check the appropriate assignee category or categories (will not be printed on the patent) : ☐ Individual ☐ Corporation or other private group entity ☐ Government

4a. The following fee(s) are submitted:
- ☒ Issue Fee
- ☐ Publication Fee (No small entity discount permitted)
- ☐ Advance Order - # of Copies

4b. Payment of Fee(s): **(Please first reapply any previously paid issue fee shown above)**
- ☐ A check is enclosed.
- ☐ Payment by credit card. Form PTO-2038 is attached.
- ☐ The director is hereby authorized to charge the required fee(s), any deficiency, or credits any overpayment, to Deposit Account Number _____ (enclose an extra copy of this form).

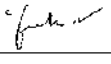
5. **Change in Entity Status** (from status indicated above)
- ☐ Applicant certifying micro entity status. See 37 CFR 1.29
- ☒ Applicant asserting small entity status. See 37 CFR 1.27
- ☐ Applicant changing to regular undiscounted fee status.

NOTE: Absent a valid certification of Micro Entity Status (see forms PTO/SB/15A and 15B), issue fee payment in the micro entity amount will not be accepted at the risk of application abandonment.

NOTE: If the application was previously under micro entity status, checking this box will be taken to be a notification of loss of entitlement to micro entity status.

NOTE: Checking this box will be taken to be a notification of loss of entitlement to small or micro entity status, as applicable.

NOTE: This form must be signed in accordance with 37 CFR 1.31 and 1.33. See 37 CFR 1.4 for signature requirements and certifications.

Authorized Signature  Date April 22, 2016

Typed or printed name H C McGurk IV Registration No. 34,964

PTOL-85 Part B (10-13) Approved for use through 10/31/2013. OMB 0651-0033 U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

RED HAT 1010
RED HAT V. CAS

Electronic Patent Application Fee Transmittal

Application Number:	14512414
Filing Date:	11-Oct-2014
Title of Invention:	MULTIPATH COMMUNICATION DEVICES AND METHODS
First Named Inventor/Applicant Name:	Eric M. Delangis
Filer:	Harold C. McGurk
Attorney Docket Number:	ED 2000

Filed as Small Entity

Filing Fees for Utility under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Utility Appl Issue Fee	2501	1	480	480

RED HAT 1010
RED HAT V. CAS

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Extension-of-Time:				
Miscellaneous:				
			Total in USD ($)	**480**

Electronic Acknowledgement Receipt	
EFS ID:	25571439
Application Number:	14512414
International Application Number:	
Confirmation Number:	9530
Title of Invention:	MULTIPATH COMMUNICATION DEVICES AND METHODS
First Named Inventor/Applicant Name:	Eric M. Delangis
Customer Number:	64828
Filer:	Harold C. McGurk
Filer Authorized By:	
Attorney Docket Number:	ED 2000
Receipt Date:	22-APR-2016
Filing Date:	11-OCT-2014
Time Stamp:	13:48:37
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	Credit Card
Payment was successfully received in RAM	$480
RAM confirmation Number	122
Deposit Account	
Authorized User	
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:	

RED HAT 1010
RED HAT V. CAS

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Issue Fee Payment (PTO-85B)	14512414_PartB_signed.pdf	2816159 21a88fd72dfa7194808cc0f191f0619cdaa6f9e6	no	1
Warnings:					
Information:					
2	Fee Worksheet (SB06)	fee-info.pdf	30062 6a23a7a5939b3c2e4395cd43e62922e128b39c55	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	2846221		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111
If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371
If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office
If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

RED HAT 1010
RED HAT V. CAS

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	ISSUE DATE	PATENT NO.	ATTORNEY DOCKET NO.	CONFIRMATION NO.
14/512,414	05/24/2016	9350649	ED 2000	9530

64828 7590 05/04/2016

Harold McGurk
The Law Office of Clay McGurk
P.O. BOX 1488
Orange, CA 92856

ISSUE NOTIFICATION

The projected patent number and issue date are specified above.

Determination of Patent Term Adjustment under 35 U.S.C. 154 (b)
(application filed on or after May 29, 2000)

The Patent Term Adjustment is 0 day(s). Any patent to issue from the above-identified application will include an indication of the adjustment on the front page.

If a Continued Prosecution Application (CPA) was filed in the above-identified application, the filing date that determines Patent Term Adjustment is the filing date of the most recent CPA.

Applicant will be able to obtain more detailed information by accessing the Patent Application Information Retrieval (PAIR) WEB site (http://pair.uspto.gov).

Any questions regarding the Patent Term Extension or Adjustment determination should be directed to the Office of Patent Legal Administration at (571)-272-7702. Questions relating to issue and publication fee payments should be directed to the Application Assistance Unit (AAU) of the Office of Data Management (ODM) at (571)-272-4200.

APPLICANT(s) (Please see PAIR WEB site http://pair.uspto.gov for additional applicants):

Eric M. Delangis, Huntsville, AL;

The United States represents the largest, most dynamic marketplace in the world and is an unparalleled location for business investment, innovation, and commercialization of new technologies. The USA offers tremendous resources and advantages for those who invest and manufacture goods here. Through SelectUSA, our nation works to encourage and facilitate business investment. To learn more about why the USA is the best country in the world to develop technology, manufacture products, and grow your business, visit SelectUSA.gov.

To: CLAYMCGURK@GMAIL.COM,,
From: PAIR_eOfficeAction@uspto.gov
Cc: PAIR_eOfficeAction@uspto.gov
Subject: Private PAIR Correspondence Notification for Customer Number 64828

May 05, 2016 05:46:56 AM

Dear PAIR Customer:

Harold McGurk
The Law Office of Clay McGurk
P.O. BOX 1488
Orange, CA 92856
UNITED STATES

The following USPTO patent application(s) associated with your Customer Number, 64828 , have new outgoing correspondence. This correspondence is now available for viewing in Private PAIR.

The official date of notification of the outgoing correspondence will be indicated on the form PTOL-90 accompanying the correspondence.

Disclaimer:
The list of documents shown below is provided as a courtesy and is not part of the official file wrapper. The content of the images shown in PAIR is the official record.

Application	Document	Mailroom Date	Attorney Docket No.
14512414	ISSUE.NTF	05/04/2016	ED 2000

To view your correspondence online or update your email addresses, please visit us anytime at https://sportal.uspto.gov/secure/myportal/privatepair.

If you have any questions, please email the Electronic Business Center (EBC) at EBC@uspto.gov with 'e-Office Action' on the subject line or call 1-866-217-9197 during the following hours:

Monday - Friday 6:00 a.m. to 12:00 a.m.

Thank you for prompt attention to this notice,

UNITED STATES PATENT AND TRADEMARK OFFICE

PATENT APPLICATION INFORMATION RETRIEVAL SYSTEM

RED HAT 1010
RED HAT V. CAS

AO 120 (Rev. 08/10)

TO:	REPORT ON THE FILING OR DETERMINATION OF AN ACTION REGARDING A PATENT OR TRADEMARK
Mail Stop 8 **Director of the U.S. Patent and Trademark Office** **P.O. Box 1450** **Alexandria, VA 22313-1450**	

In Compliance with 35 U.S.C. § 290 and/or 15 U.S.C. § 1116 you are hereby advised that a court action has been filed in the U.S. District Court Eastern District of Texas on the following

☐ Trademarks or ☑ Patents. (☐ the patent action involves 35 U.S.C. § 292.):

DOCKET NO.	DATE FILED	U.S. DISTRICT COURT
2:22-cv-287	7/27/2022	Eastern District of Texas

PLAINTIFF	DEFENDANT
Competitive Access Systems, Inc.	Oracle Corporation and Oracle America, Inc.

	PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1	8,228,801	7/24/2012	Competitive Access Systems, Inc.
2	9,350,649	5/24/2016	Competitive Access Systems, Inc.
3	10,868,908	12/15/2020	Competitive Access Systems, Inc.
4			
5			

In the above entitled case, the following patent(s)/ trademark(s) have been included:

DATE INCLUDED	INCLUDED BY ☐ Amendment ☐ Answer ☐ Cross Bill ☐ Other Pleading	
PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1		
2		
3		
4		
5		

In the above entitled case, the following decision has been rendered or judgement issued:

DECISION/JUDGEMENT

CLERK	(BY) DEPUTY CLERK	DATE

Copy 1—Upon initiation of action, mail this copy to Director **Copy 3—Upon termination of action, mail this copy to Director**
Copy 2—Upon filing document adding patent(s), mail this copy to Director **Copy 4—Case file copy**

RED HAT 1010
RED HAT V. CAS

AO 120 (Rev. 08/10)

TO:	REPORT ON THE FILING OR DETERMINATION OF AN ACTION REGARDING A PATENT OR TRADEMARK
Mail Stop 8 **Director of the U.S. Patent and Trademark Office** **P.O. Box 1450** **Alexandria, VA 22313-1450**	

In Compliance with 35 U.S.C. § 290 and/or 15 U.S.C. § 1116 you are hereby advised that a court action has been filed in the U.S. District Court Western Districtof Texas on the following

☐ Trademarks or ☑ Patents. (☐ the patent action involves 35 U.S.C. § 292.):

DOCKET NO.	DATE FILED	U.S. DISTRICT COURT
1:23-cv-286	3/16/2022	Western District of Texas

PLAINTIFF	DEFENDANT
Competitive Access Systems, Inc.	Oracle Corporation and Oracle America, Inc.

	PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1	8,228,801	7/24/2012	Competitive Access Systems, Inc.
2	9,350,649	5/24/2016	Competitive Access Systems, Inc.
3	10,868,908	12/15/2020	Competitive Access Systems, Inc.
4			
5			

In the above—entitled case, the following patent(s)/ trademark(s) have been included:

DATE INCLUDED	INCLUDED BY ☐ Amendment ☐ Answer ☐ Cross Bill ☐ Other Pleading	
PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1		
2		
3		
4		
5		

In the above—entitled case, the following decision has been rendered or judgement issued:

DECISION/JUDGEMENT

CLERK	(BY) DEPUTY CLERK	DATE
PHILIP J. DEVLIN		03/20/2023

Copy 1—Upon initiation of action, mail this copy to Director **Copy 3—Upon termination of action, mail this copy to Director**
Copy 2—Upon filing document adding patent(s), mail this copy to Director **Copy 4—Case file copy**

RED HAT 1010
RED HAT V. CAS

AO 120 (Rev. 08/10)

TO:	Mail Stop 8 Director of the U.S. Patent and Trademark Office P.O. Box 1450 Alexandria, VA 22313-1450	REPORT ON THE FILING OR DETERMINATION OF AN ACTION REGARDING A PATENT OR TRADEMARK

In Compliance with 35 U.S.C. § 290 and/or 15 U.S.C. § 1116 you are hereby advised that a court action has been filed in the U.S. District Court for the District of Arizona on the following

☑ Trademarks or ☐ Patents. (☐ the patent action involves 35 U.S.C. § 292.):

DOCKET NO.	DATE FILED	U.S. DISTRICT COURT
2:23-cv-00368-DWL	3/1/2023	for the District of Arizona

PLAINTIFF	DEFENDANT
Christian Copyright Licensing International, LLC	Multitracks.com LLC

	PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1	U.S. Reg. No. 4,872,	12/15/2015	Multitracks.com LLC
2			
3			
4			
5			

In the above—entitled case, the following patent(s)/ trademark(s) have been included:

DATE INCLUDED	INCLUDED BY ☐ Amendment ☐ Answer ☐ Cross Bill ☐ Other Pleading	
PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1		
2		
3		
4		
5		

In the above—entitled case, the following decision has been rendered or judgement issued:

DECISION/JUDGEMENT

CLERK	(BY) DEPUTY CLERK	DATE
DEBRA LUCAS		3/10/23

Copy 1—Upon initiation of action, mail this copy to Director **Copy 3—Upon termination of action, mail this copy to Director**
Copy 2—Upon filing document adding patent(s), mail this copy to Director **Copy 4—Case file copy**

RED HAT 1010
RED HAT V. CAS

Patent Trial and Appeal Board

A petition has been filed in Patent Number 9,350,649, Application Number 14/512,414 on May 8, 2023.

The AIA Review Number is IPR2023-00880.

To view the documents filed in this petition, go to https://ptab.uspto.gov and Search for the AIA Review Number.

- Enter your search criteria on the "Search PTAB" page
- Type in the AIA Review Number or Patent Number
- You will need to answer the CAPTCHA to prove that you are not a robot.
- Click on the "Search" button
- The search results will appear identifying the AIA Review Number
- Click on the "View Documents" button
- A pop up window will appear with a list of documents
- Click on the "Download" button to download the document.

Questions regarding this notice should be directed to the Patent Trial and Appeal Board at 571-272-7822.

Trials@uspto.gov
571-272-7822

Paper No. 3

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

ORACLE CORPORATION,
Petitioner,

v.

COMPETITIVE ACCESS SYSTEMS, INC.,
Patent Owner.

IPR2023-00880
Patent 9,350,649 B2

Mailed: May 23, 2023

Before EMANUEL MINES, *Trial Paralegal*

NOTICE OF FILING DATE ACCORDED TO PETITION
AND
TIME FOR FILING PATENT OWNER PRELIMINARY RESPONSE

The petition for *inter partes* review in the above proceeding has been accorded the filing date of May 8, 2023.

RED HAT 1010
RED HAT V. CAS

Patent Owner may file a preliminary response to the petition no later than three months from the date of this notice. The preliminary response is limited to setting forth the reasons why the requested review should not be instituted. Patent Owner may also file an election to waive the preliminary response to expedite the proceeding. For more information, please consult the Office Patent Trial Practice Guide, 77 Fed. Reg. 48756 (Aug. 14, 2012), which is available on the Board Web site at http://www.uspto.gov/PTAB.

Patent Owner is advised of the requirement to submit mandatory notice information under 37 C.F.R. § 42.8(a)(2) within 21 days of service of the petition.

The parties are encouraged to use the heading on the first page of this Notice for all future filings in the proceeding.

The parties are advised that under 37 C.F.R. § 42.10(c), recognition of counsel *pro hac vice* requires a showing of good cause. The parties are authorized to file motions for *pro hac vice* admission under 37 C.F.R. § 42.10(c). Such motions shall be filed in accordance with the "Order -- Authorizing Motion *for Pro Hac Vice* Admission" in Case IPR2013-00639, Paper 7, a copy of which is available on the Board Web site under "Representative Orders, Decisions, and Notices."

The parties are reminded that unless otherwise permitted by 37 C.F.R. § 42.6(b)(2), all filings in this proceeding must be made electronically in Patent Trial and Appeal Board End to End (PTAB E2E), accessible from the Board Web site at http://www.uspto.gov/PTAB. To file documents, users must register with PTAB E2E. Information regarding how to register with and use PTAB E2E is available at the Board Web site.

If there are any questions pertaining to this notice, please contact the Patent Trial and Appeal Board at 571-272-7822.

For PETITIONER:

Harper Batts
Chris Ponder
Jeffrey Liang
SHEPPARD, MULLIN, RICHTER & HAMPTON LLP
hbatts@sheppardmullin.com
cponder@sheppardmullin.com
jliang@sheppardmullin.com

For PATENT OWNER:

Harold McGurk
The Law Office Of Clay McGurk
PO Box 1488
Orange, CA 92856
claymcgurk@gmail.com

NOTICE CONCERNING ALTERNATIVE DISPUTE RESOLUTION (ADR)

The Patent Trial and Appeal Board (PTAB) strongly encourages parties who are considering settlement to consider alternative dispute resolution as a means of settling the issues that may be raised in an AIA trial proceeding. Many AIA trials are settled prior to a Final Written Decision. Those considering settlement may wish to consider alternative dispute resolution techniques early in a proceeding to produce a quicker, mutually agreeable resolution of a dispute or to at least narrow the scope of matters in dispute. Alternative dispute resolution has the potential to save parties time and money.

Many non-profit organizations, both inside and outside the intellectual property field, offer alternative dispute resolution services. Listed below are the names and addresses of several such organizations. The listings are provided for the convenience of parties involved in cases before the PTAB; the PTAB does not sponsor or endorse any particular organization's alternative dispute resolution services. In addition, consideration may be given to utilizing independent alternative dispute resolution firms. Such firms may be located through a standard keyword Internet search.

CPR INSTITUTE FOR DISPUTE RESOLUTION	**AMERICAN INTELLECTUAL PROPERTY LAW ASSOCIATION (AIPLA)**	**AMERICAN ARBITRATION ASSOCIATION (AAA)**	**WORLD INTELLECTUAL PROPERTY ORGANIZATION (WIPO)**	**AMERICAN BAR ASSOCIATION (ABA)**
Telephone: (212) 949-6490	Telephone: (703) 415-0780	Telephone: (212) 484-3266	Telephone: 41 22 338 9111	Telephone : (202) 662-1000
Fax: (212) 949-8859	Fax: (703) 415-0786	Fax: (212) 307-4387	Fax: 41 22 733 5428	N/A
575 Lexington Ave	241 18th Street, South, Suite 700	140 West 51st Street	34, chemin des Colombettes	1050 Connecticut Ave, NW
New York, NY 10022	Arlington, VA 22202	New York, NY 10020	CH-1211 Geneva 20, Switzerland	Washington D.C. 20036
www.cpradr.org	www.aipla.org	www.adr.org	www.wipo.int	www.americanbar.org

If parties to an AIA trial proceeding consider using alternative dispute resolution, the PTAB would like to know whether the parties ultimately decided to engage in alternative dispute resolution and the reasons why or why not. If the parties actually engage in alternative dispute resolution, the PTAB would be interested to learn what mechanism (e.g., arbitration, mediation, etc.) was used and the general result. Such a statement from the parties is not

required but would be helpful to the PTAB in assessing the value of alternative dispute resolution to parties involved in AIA trial proceedings. To report an experience with ADR, please forward a summary of the particulars to the following email address: PTAB_ADR_Comments@uspto.gov

Trials@uspto.gov
571-272-7822

Paper 11
Date: October 30, 2023

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

ORACLE CORPORATION,
Petitioner,

v.

COMPETITIVE ACCESS SYSTEMS, INC.,
Patent Owner.

IPR2023-00880
Patent 9,350,649 B2

Before BARBARA A. PARVIS, STEVEN M. AMUNDSON, and JASON M. REPKO, *Administrative Patent Judges.*

PARVIS, *Administrative Patent Judge.*

DECISION
Settlement Prior to Institution of Trial
Granting-In-Part and *Denying-In-Part* Joint Request to Treat Settlement Agreement as Business Confidential Information
37 C.F.R. § 42.74

RED HAT 1010
RED HAT V. CAS

I. INTRODUCTION

On October 24, 2023, with the Board's authorization, Petitioner Oracle Corporation and Patent Owner Competitive Access Systems, Inc. (collectively referred to as "the Parties") filed a Joint Motion to Terminate or Dismiss the above-identified proceeding due to settlement. Paper 10 ("Joint Motion"). In support of the Joint Motion, the Parties filed a Settlement Agreement (Ex. 1030 ("Settlement Agreement")), as well as a Joint Request to Keep Separate the Settlement Agreement pursuant to 35 U.S.C. § 317(b) and 37 C.F.R. § 42.74(c) (Paper 9 ("Joint Request")).

II. DISCUSSION

In the Joint Motion, the Parties represent that they "have settled with respect to the challenged patent and have reached agreement to terminate this IPR." Paper 10, 2. The Parties further represent that they "have filed a stipulation dismissing the parallel district court action with prejudice." *Id.* at 4 (citing *Competitive Access Systems, Inc. v. Oracle Corp.*, No. 1:23-cv-00286 (W.D. Tex.) (October 20, 2023) (ECF No. 86)). The Parties also represent that they "are unaware of any other IPR that is challenging" the above-identified patent at issue. *Id.* The Parties represent that "the written agreement that contemplates the termination of this proceeding has been filed with this motion as Exhibit 1030,"[1] and "certify that there are no other agreements, oral or written, between the parties made in connection with, or in contemplation of, the termination of the proceeding." *Id.*

This proceeding is at an early stage, and we have not yet decided whether to institute a trial in the proceeding. In view of the early stage of

[1] The Parties state that Ex. 1030 is "a true copy of the parties' written agreement that contemplates the termination of this proceeding." Paper 9, 2.

RED HAT 1010
RED HAT V. CAS

the proceeding and the settlement between the Parties, we determine that good cause exists to terminate the proceeding with respect to the Parties.

The Parties also filed a Joint Request that the Settlement Agreement be treated as business confidential information and be kept separate from the file of the patent involved in this *inter partes* proceeding. Paper 9, 2. After reviewing the Settlement Agreement between Petitioner and Patent Owner, we find that the Settlement Agreement contains confidential business information regarding the terms of settlement. We determine that good cause exists to treat the Settlement Agreement between Petitioner and Patent Owner as business confidential information, to keep it separate from the file of the involved patent, and to limit its availability pursuant to 35 U.S.C. § 317(b) and 37 C.F.R. § 42.74(c).

In the Joint Request, the parties "jointly request that the Board order that in the event a person or entity makes a written request, as stated in 37 C.F.R. § 42.74(c)(1)-(2), for access to the settlement agreement, that any such written request be served upon the parties on the day the written request is provided to the Board." Paper 9, 1. Our rules do not provide for any written request to the parties, and the parties have provided no basis for doing so. Thus, we deny this part of the parties' request.

This Order does not constitute a final written decision pursuant to 35 U.S.C. § 318(a).

III. ORDER

Accordingly, for the reasons discussed above, it is:

ORDERED that the Joint Motion to Terminate (Paper 10) is *granted*;

FURTHER ORDERED that the Petition (Paper 2) is *dismissed*;

FURTHER ORDERED that IPR2023-00880 is *terminated*; and

FURTHER ORDERED that the Joint Request to Keep Separate

(Paper 9) is *granted-in-part* and *denied-in-part*, and the Settlement Agreement shall be kept separate from the file of Patent 9,350,649 B2 and made available only to Federal Government agencies on written request, or to any person on a showing of good cause, pursuant to 35 U.S.C. § 317(b) and 37 C.F.R. § 42.74(c), but no written request will be served upon the parties.

For PETITIONER:

Harper Batts
Chris Ponder
Jeffrey Liang
SHEPPARD, MULLIN, RICHTER & HAMPTON LLP
hbatts@sheppardmullin.com
cponder@sheppardmullin.com
jliang@sheppardmullin.com

For PATENT OWNER:

Jason A. Engel
Katherine L. Allor
Samuel P. Richey
Nolan R. Hubbard
K&L GATES LLP
Jason.Engel.PTAB@klgates.com
Katy.Allor@klgates.com
Samuel.Richey@klgates.com
Nolan.Hubbard@klgates.com