5th IEEE International Symposium on

Network Computing and Applications

NCA 06

24–26 July 2006

Cambridge, Massachusetts, USA

Sponsored by
IEEE Computer Society Technical Committee on Distributed Processing

IEEE COMPUTER SOCIETY

IEEE

PAN Exhibit 1022
Page 1 of 16

Proceedings

Fifth IEEE International Symposium on Network Computing and Applications

---

NCA 2006

Proceedings

Fifth IEEE International Symposium on Network Computing and Applications

(NCA 2006)

24–26 July 2006

Cambridge, Massachusetts

Los Alamitos, California
Washington • Tokyo

Copyright © 2006 by The Institute of Electrical and Electronics Engineers, Inc.
All rights reserved.

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries may photocopy beyond the limits of US copyright law, for private use of patrons, those articles in this volume that carry a code at the bottom of the first page, provided that the per-copy fee indicated in the code is paid through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

Other copying, reprint, or republication requests should be addressed to: IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, P.O. Box 133, Piscataway, NJ 08855-1331.

The papers in this book comprise the proceedings of the meeting mentioned on the cover and title page. They reflect the authors' opinions and, in the interests of timely dissemination, are published as presented and without change. Their inclusion in this publication does not necessarily constitute endorsement by the editors, the IEEE Computer Society, or the Institute of Electrical and Electronics Engineers, Inc.

IEEE Computer Society Order Number P2640
ISBN-13: 978-0-7695-2640-9
ISBN-10: 0-7695-2640-3
Library of Congress Control Number 2006928423

Additional copies may be ordered from:

IEEE Computer Society
Customer Service Center
10662 Los Vaqueros Circle
P.O. Box 3014
Los Alamitos, CA 90720-1314
Tel: + 1 800 272 6657
Fax: + 1 714 821 4641
http://computer.org/cspress
csbooks@computer.org

IEEE Service Center
445 Hoes Lane
P.O. Box 1331
Piscataway, NJ 08855-1331
Tel: + 1 732 981 0060
Fax: + 1 732 981 9667
http://shop.ieee.org/store/
customer-service@ieee.org

IEEE Computer Society
Asia/Pacific Office
Watanabe Bldg., 1-4-2
Minami-Aoyama
Minato-ku, Tokyo 107-0062
JAPAN
Tel: + 81 3 3408 3118
Fax: + 81 3 3408 3553
tokyo.ofc@computer.org

Individual paper REPRINTS may be ordered at: reprints@computer.org

Editorial production by Randall S. Bilof
Cover art production by Joe Daigle/Studio Productions
Printed in the United States of America by The Printing House

IEEE COMPUTER SOCIETY

IEEE

IEEE Computer Society
Conference Publishing Services
http://www.computer.org/proceedings/

Contents

Proceedings of the Fifth IEEE International Symposium on Network Computing and Applications

NCA 2006

Message from the Program Committee Co-Chairs and Steering Committee Chair ix
Organizing Committee x
Program Committee xi
Trustworthy Network Computing Workshop Program Committee xii
Adaptive Grid Computing Workshop Program Committee xii
External Reviewers xii

Keynote Speech Abstracts

Experience with Some Principles for Building an Internet-Scale Reliable System 3
Mike Afergan
Challenges in Managing Complexity: Autonomic Computing 4
Alan Ganek
The +CAL Algorithm Language 5
Leslie Lamport
Application-Aware Reliability and Security: The Trusted ILLIAC Approach 6
Ravishankar K. Iyer

Networking Protocols

A Tree-Turn Model for Irregular Networks 11
Jiazheng Zhou, Xuan-Yi Lin, and Yeh-Ching Chung
Experience Using Active and Passive Mapping for Network Situational Awareness 19
Seth Webster, Richard Lippmann, and Marc Zissman
Multi-Path Multi-Channel Routing Protocol 27
Bo Yan and Hamid Gharavi
JetStream: Achieving Predictable Gossip Dissemination by Leveraging Social Network Principles 32
Jay A. Patel, Indranil Gupta, and Noshir Contractor
Self* Architecture for Trajectory Tracking in Wireless Sensor Networks 40
F. Claerhout, A. K. Datta, M. Gradinariu, and M. Hurfin
Parallelization of Geant4 Using TOP-C and Marshalgen 48
Gene Cooperman, Viet Ha Nguyen, and Igor Malioutov

PAN Exhibit 1022
Page 5 of 16

Software Aging and Rejuvenation in a SOAP-Based Server 56
Luis Silva, Henrique Madeira, and João Gabriel Silva

Utilizing Network Processors in Distributed Enterprise Environments 66
Paul Royal, Mitch Halpin, Ada Gavrilovska, and Karsten Schwan

Peer-to-Peer and Publish Subscribe Systems

Faster Event Forwarding in a Content-Based Publish-Subscribe System through Lookup Reuse 77
Zhenhui Shen and Srikanta Tirthapura

Improved Result Ranking in P2P File-Sharing Systems by Probing for Metadata 85
Wai Gen Yee, Linh Thai Nguyen, and Ophir Frieder

Identifying Known and Unknown Peer-to-Peer Traffic 93
Fivos Constantinou and Panayiotis Mavrommatis

Real-Time Scheduling

Video-on-Demand Equipment Allocation 103
Frederic Thouin, Mark Coates, and Dominic Goodwill

Slotted Optical Burst Switching (SOBS) Networks 111
Zhenghao Zhang, Lin Liu, and Yuanyuan Yang

Implementing the Advanced Switching Minimum Bandwidth Egress Link Scheduler 118
Raúl Martínez, Francisco J. Alfaro, and José L. Sánchez

ContagAlert: Using Contagion Theory for Adaptive, Distributed Alert Propagation 126
Michael Treaster, William Conner, Indranil Gupta, and Klara Nahrstedt

Security and Trust

A Proxy Network Topology and Reconfiguration Algorithm for Resistance to DoS Attacks 137
Stephen Frechette and D. R. Avresky

SecLEACH—A Random Key Distribution Solution for Securing Clustered Sensor Networks 145
Leonardo B. Oliveira, Hao C. Wong, M. Bern, Ricardo Dahab, and A. A. F. Loureiro

Middleware

Providing e-Transaction Guarantees in Asynchronous Systems with Inaccurate Failure Detection 155
Paolo Romano and Francesco Quaglia

From DC-Nets to pMIXes: Multiple Variants for Anonymous Communications 163
Carlos Aguilar Melchor and Yves Deswarte

Distributed Storage

Optimizing Cauchy Reed-Solomon Codes for Fault-Tolerant Network Storage Applications 173
James S. Plank and Lihao Xu

Network-Aware Adaptation Techniques for Mobile File Systems 181
Benjamin Atkin and Kenneth P. Birman

PAN Exhibit 1022
Page 6 of 16

Congestion Control for Distributed Hash Tables 189
Fabius Klemm, Jean-Yves Le Boudec, and Karl Aberer

Comparison of Redundancy Schemes for Distributed Storage Systems 196
Peter Sobe and Kathrin Peter

A Primary-Backup Protocol for In-Memory Database Replication 204
Lásaro Camargos, Fernando Pedone, and Rodrigo Schmidt

A Flexible Replication Mechanism with Extended Database Connection Layers 212
Nobutatsu Nakamura, Ken'ichiro Fujiyama, Eiji Kawai, and Hideki Sunahara

Short Papers

Analysis Models for Blind Search in Unstructured Overlays 223
Bin Wu and Ajay D. Kshemkalyani

A Novel Decision Strategy of Vertical Handoff in Overlay Wireless Networks 227
Wei Shen and Qing-An Zeng

A System for Secure IP Telephone Conferences 231
Axel Treßel and Jörg Keller

A Distributed Architecture for Gracefully Degradable Web-Based Services 235
Mauro Andreolini, Sara Casolari, and Michele Colajanni

Full QoS Support with 2 VCs for Single-Chip Switches 239
Alejandro Martínez, Francisco J. Alfaro, José L. Sánchez, and José Duato

A Flexible Cross Connect LCAS for Bandwidth Maximization in 2.5G EoS 243
Hou-Cheng Tzeng and Ching-Te Chiu

Decentralized Load Balancing on Unstructured Peer-2-Peer Computing Grids 247
Jie Hu and Raymond Klefstad

Finding Shared Fragments in Large Collections of Web Pages for Fragment-Based Web Caching 251
Junchang Ma and Zhimin Gu

Lossy Node Identification in Wireless Sensor Network 255
Yongjun Li, Wandong Cai, Wei Wang, and Guangli Tian

IntranetWeb: A Peer-to-Peer Based Scalable Web Caching System 259
Huifang Cheng and Zhimin Gu

A Novel Probability Evaluation Method for Selective Forwarding Routing in Wireless Sensor Network 263
Li-jia Chen, Jing Wu, Pu-liu Yan, Jian-guo Zhou, and Hao Jiang

Adaptive Grid Computing Workshop

A Multipolicy Authorization Framework for Grid Security 269
Bo Lang, Ian Foster, Frank Siebenlist, Rachana Ananthakrishnan, and Tim Freeman

An Adaptive Quality of Service Based Scheduling Algorithm for Wide Area Large Scale Problems 273
Wilson Lozano and Wilson Rivera

Design and Implementation of a Plain Grid Monitoring and Information Service 277
Weibin Pei, Jun Wang, Pengzhi Liu, Zhongliang Chen, Chunhao Feng, and Zhi Wang

PAN Exhibit 1022
Page 7 of 16

Trustworthy Network Computing Workshop

Trust Assessment from Observed Behavior: Toward an Essential Service for Trusted Network Computing 285
Partha Pal, Franklin Webber, Michael Atighetchi, and Nate Combs

An Improved Test Derivation Technique 293
Harry Gros-Desormeaux, Hacène Fouchal, and Philippe Hunel

Author Index 301

PAN Exhibit 1022
Page 8 of 16

Experience Using Active and Passive Mapping for Network Situational Awareness*

Seth Webster Dr. Richard Lippmann Dr. Marc Zissman
MIT Lincoln Laboratory
244 Wood St.
Lexington, MA 02420
{*swebster,lippmann,maz*}*@ll.mit.edu*

Abstract

Passive network mapping has often been proposed as an approach to maintain up-to-date information on networks between active scans. This paper presents a comparison of active and passive mapping on an operational network. On this network, active and passive tools found largely disjoint sets of services and the passive system took weeks to discover the last 15% of active services. Active and passive mapping tools provided different, not complimentary information. Deploying passive mapping on an enterprise network does not reduce the need for timely active scans due to non-overlapping coverage and potentially long discovery times.

1. Introduction

An important requirement for maintaining network security is to have up-to-date network situational awareness concerning the hosts in a network and the TCP and UDP services they run. Active network scanning, as is provided by the open-source tools Nmap [1] and Nessus [2], has historically been used to obtain this information. These tools actively send probes over a network to discover hosts, open ports, and operating systems. Active scanning has the benefit of proactively contacting all hosts on a network to determine their status, but also has many disadvantages. These include long scan times for large networks, the possibility of crashing services or hosts, a lack of visibility into subnets protected by firewalls that block scans, a tendency to falsely trigger intrusion detection systems, and the need to provide credentials on some probed hosts for a complete host analysis. Recently, passive mapping tools that analyze network traffic to identify hosts, open ports, and operating systems have been suggested to augment active scanning. Open-source tools, such as Siphon [3], P0f [4], and Ettercap [5], have been available for some time and commercial products, such as Tennable's NeVO [6] and Sourcefire's RNA [7], have appeared in the last few years. Often-cited benefits of passive mapping are that hosts and network devices can not be harmed, network infrastructure is not impacted, host and service information is continuously updated, and information is provided even for networks behind firewalls where active scans are not permitted. Generally recognized limitations are that passive mapping can not find unused services and is difficult to deploy on switched networks where there are no natural choke points for traffic monitoring.

Despite the increased interest in passive mapping, little has been published concerning its performance. The goal of the research described in this paper was to compare active and passive mapping on an actual network to better understand the strengths and limitations of these two approaches. Metrics of interest include the overlap in devices and services discovered by active and passive mapping, the time required for passive mapping to discover services, and the the average rate at which the network changes.

This study is limited because we analyze data at only a single site and only over three months. Traffic between an enterprise network and the internet differs dramatically across sites and time [8] and the topology of enterprise networks differ. Our results should not be taken as being representative of networks substantially different from the one analyzed. One important characteristics of this network is the use of firewalls to isolate network segments. A border firewall is used to protect all hosts in the enterprise and internal firewalls are used to isolate internal intranet subnets. These firewalls break up the network and limit the visibility of active scanning and passive mapping tools. A second important characteristic of this network is the type and volume of traffic generated. which affects how rapidly passive traf-

*This work was sponsored by the Federal Aviation Administration under Air Force Contract F19628-00-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the United States Government.

Fifth IEEE International Symposium on Network Computing and Applications (NCA'06)
0-7695-2640-3/06/$20.00 © 2006 IEEE
Authorized licensed use limited to: Klarquist Sparkman LLP. Downloaded on March 31,2026 at 20:01:57 UTC from IEEE Xplore. Restrictions apply.

fic analysis discovers devices and ports. In addition, security policies prohibit use of peer-to-peer file sharing, which can be a large component of ISP traffic [9].

Major contributions of this paper are (1) It provides the first comparison we are aware of between active and passive approaches to network mapping on an operational network, (2) It demonstrates that passive and active scanning may have significantly different views into the network, (3) Passive discovery times for different TCP/IP services are measured and range from seconds to weeks, and (4) The rate of change across the analyzed network is measured and shown to be at least one new or removed host or service every one to two days. The remainder of this paper reviews the current literature on passive mapping techniques, describes in detail the network and scanning tools analyzed, gives details on the data collected from the tools, and presents the results of the various analyses.

2. Background

A number of commercial and open source network scanning tools are available to administrators. These tools discover and monitor hosts in a network by sending ICMP, TCP, UDP, and ARP packets. A simple open-source tool named Nmap [1] is often used to discover hosts, their operating systems, and open ports. Other, more advanced active scanners, such as the open-source Nessus vulnerability scanner [2] find hosts and determine whether they contain known software vulnerabilities. Many system administrators run Nessus or other vulnerability scanners periodically to identify hosts that might be compromised by remote attackers. These tools are widely used and generally accepted as accurate. They, however, have a few major limitations. As noted above, when scanner traffic is blocked by firewalls, protected hosts will not be seen. Scanning can also cause intrusion detection systems to issue false alarms, and some types of scans can cause programs and hosts to crash. Our experience with vulnerability scanners, even when using "safe" scans, is that they can crash printers and custom network servers. Scanning can also be slow. A scan of all 65,535 TCP ports and limited UDP ports on all hosts in the network described in this paper using Nessus takes roughly a week. This long scan time makes it impossible to monitor rapid network changes.

A passive network mapping system attempts to discover hosts and open TCP/UDP ports on a network by passively analyzing network traffic. The open-source tool Ntop [10] was one of the first widely available passive network mapping tools. It analyses captured traffic and creates web pages that present a wide range of information on the most active protocols and hosts. Other more limited open-source passive reconnaissance tools are available that identify hosts, open ports, and operating systems. Currently maintained tools that are in common use include Ettercap [5] and P0f [4]. Commercial companies such as Tenable [6] and Sourcefire [7] provide passive analysis tools that can be used on enterprise networks and include management and reporting capabilities not found in the above open-source tools.

Commercial passive mapping systems are often marketed for their ability to continuously monitor networks without damaging hosts, provide more timely notification of new hosts and services than active scanning, and gain visibility into networks behind firewalls. Little, however, has been published that verifies these claims or explores limitations of these tools. Dayioglu [11] discusses the possibility of using a passive mapping system in conjunction with an intrusion detection system to decrease the false alarm rate. Lippmann [12] demonstrates that the accuracy of operating system identification for Ettercap and P0f is low and describes approaches to improve performance. Kuntzelman [13], in a master's thesis, compares active and passive mapping systems but in a small flat testbed network with simulated clients and servers. Because the performance of active and passive mapping systems is strongly governed by network topology and traffic patterns, this analysis does not provide any insight into real-world performance. Finally, Montigny [14] describes some new approaches to identify host operating systems, TCP/UDP services, and network infrastructure devices with passive analysis, but provides no evaluation of these approaches with real traffic.

3. Network Topology

Analyses presented in this paper were performed on a large DMZ of an enterprise network. This DMZ network contains roughly 800 workstations and servers that were positioned outside a strong internal firewall to allow public access to servers and collaboration with other sites. As shown in Figure 1, this DMZ is located between an external firewall that leads to the internet and an internal firewall that leads to a large intranet. The DMZ network contains hosts and subnets that are directly connected to a backbone switch and other subnets that are protected behind firewalls, including an administrative subnet. Inbound and outbound Internet traffic averaged 53 TCP connections per second over the three months studied. The peak rate on this link was 10-20 Mb/sec and the average rate was roughly 1 Mb/sec. There were 800 hosts on the network, and many of the connections between these and outside hosts support collaborative data analysis. Active scans are performed from the administrative subnet and the firewall on this subnet is configured to pass scan traffic.

Security for DMZ hosts is provided by both a border router and an external firewall. The border router blocks traffic for unwanted protocols and services such as Net-

PAN Exhibit 1022
Page 10 of 16

Authorized licensed use limited to: Klarquist Sparkman LLP. Downloaded on March 31,2026 at 20:01:57 UTC from IEEE Xplore. Restrictions apply.

BIOS. The external firewall blocks all traffic not originating from or destined to specific IP addresses and ports. DMZ hosts can only initiate outbound connections to specific ports and/or IP addresses and servers will only receive connections from the Internet on specific ports. Firewalls on DMZ subnets in Figure 1 prevent access from other DMZ hosts or subnets. Such multiple firewalls and subnets are found in many enterprise networks. Often, firewalls protecting subnets are locally administered while peripheral border firewalls are centrally administered. Multiple administrative domains, such as these, frequently make it difficult to obtain permission to scan all subnets of a network.

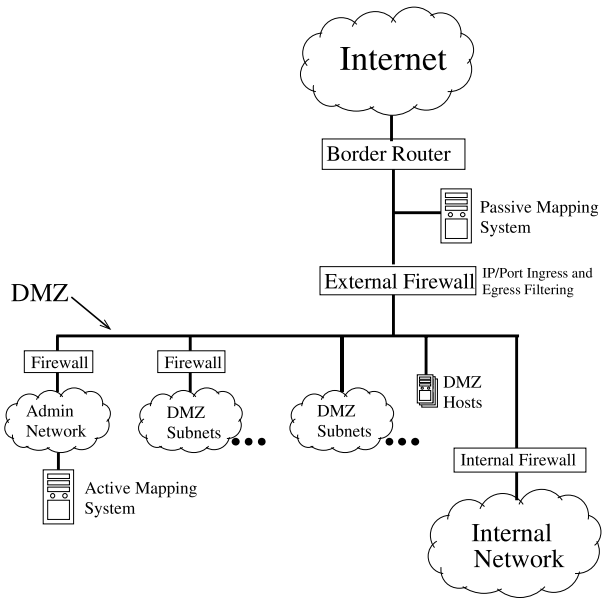


Figure 1. Network topology

4. Active and Passive Mapping Tools Used in This Study

Nessus was used to perform the active scanning for this study because it was the tool already in use by local network administrators. It was configured to scan all possible IP addresses and then scan all TCP ports on any address found to be active. Because of the time required and the uncertain nature of UDP scanning, UDP scans were only performed on ports that were likely to be open. For example, if the TCP port in a common TCP/UDP port pair was discovered, the scanner would attempt to find the corresponding UDP port on the same IP address.

The passive mapping system, LANscape, was designed to monitor multiple remote networks simultaneously, eliminate false detections of services caused by transient use of high-numbered ports, and require limited memory and processing resources. It has been used successfully for over a year at nine government sites. LANscape monitors both TCP and UDP network traffic and builds a catalog of all IP addresses seen and all TCP and UDP ports open on those systems.

Open TCP ports are found by looking for SYN-ACK packets. Because these packets are not useful for network reconnaissance, they are very rarely sent unsolicited and provide a robust way to identify true server responses with little overhead. Because UDP traffic lacks an inherent notion of a connection, it is more ambiguous. All UDP packets are monitored and sorted into connections. A client is assumed to be the sender of the first UDP packet seen. Occasionally, the first packet in a connection will be dropped by the sniffer, resulting in the system considering the source port to be a potentially open server port. These mistakes are generally filtered out by the transient port detection algorithm.

LANscape dismisses high-numbered transient ports, such as those used by FTP, to prevent identifying them as permanent servers. Connections for ports above 512 are placed in a per-host queue of length seven. Ports that accumulate three or more connections from at least two different clients are dequeued and made permanent. Ports pushed off the back of the queue by other potentially transient ports before accumulating enough connections are discarded. A conservative estimate based on statistics from ports below 512 is that the number of accidentally rejected permanent ports above 512 is less than 3% of the total number of ports. These rejected ports may slightly alter the results of the passive-active overlap analysis, but do not affect the network change or time to discovery analyses.

5. Data Collection

Data was collected on the network described in Section 3 for 86 days in 2004. During this time, two active scans were performed, the first ending on the first day and the second on the 85th day. Passive mapping data was collected from the 10th day until the 86th day with the exception of the 13 days between the 59th and 72nd days due to the passive mapping system being down. Figure 2 shows graphically days where data was collected. Active scans were performed by the network security personnel as part of routine monitoring and their timing was not optimal for this study. In particular, data would have been easier to analyze if the initial active scan coincided with the beginning of passive monitoring.

PAN Exhibit 1022
Page 11 of 16

Authorized licensed use limited to: Klarquist Sparkman LLP. Downloaded on March 31,2026 at 20:01:57 UTC from IEEE Xplore. Restrictions apply.

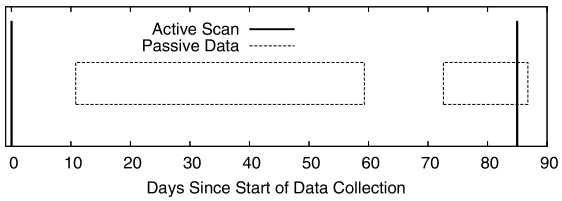


Figure 2. Dates of data collection

6. Results

6.1. Developing Ground Truth and Data Analysis

Because the DMZ network was large, dynamic, and administered by different organizations, it was not possible to obtain ground truth concerning hosts and services on the network. To overcome the lack of truth, results from one mapping technique served as a baseline to evaluate the other. For example, services found in both the first and last active scan are assumed to have been on the network for all times between the scans. This subset of known active services is then used to evaluate the performance of passive mapping.

Using the passive system to determine when a given service was on the network is more complicated. Based on the traffic statistics reported every 30 seconds, the number and distribution of successful connections to a given service is known with relatively high fidelity. However, because the passive system does not record failed connection attempts, we do not know if a lack of successful connections over a long time interval is a result of the service simply not being used or a result of the service being removed from the network. Examining the distribution of successful connection inter-arrival times, however, makes it possible to create an estimate of the longest of gap between connections that that is likely to be seen for each service. Gaps longer than this threshold indicate that a service was not present on the network to answer incoming connections. Figure 3 shows plots of four different inter-arrival time distributions for four IP/port pairs. These figures are plotted on a log-log scale and, because the passive system aggregates connection statistics in 30 second intervals, all inter-arrival values less than 30 seconds are extrapolated from the number of connections reported in that interval. These graphs show the variety of distributions found for various services. In general, the following four types of distributions were seen:

- A relatively slow falloff, as seen in Figure 3(a)
- A much steeper falloff but a long tail, as seen in Figure 3(b)
- A periodic distribution dominated by evenly spaced peaks, as seen in Figure 3(c).
- A centrally peaked distribution (in log-log space), as seen in Figure 3(d).

These plots show that different services have different inter-arrival time distributions and many of these distributions have long tails. Similar results were obtained by Crovella and Bestavros who demonstrated that web traffic inter-arrival times can be modeled with a long-tailed Pareto distribution [15]. Based on this and previous analyses, it was decided to set a gap threshold to identify terminated services to be $4 \times maxgap$ where $maxgap$ is the largest inter-arrival time seen for a service. In addition, these distributions suggest that passive traffic analysis may have to sample data for long time intervals to discover all services because extremely long inter-arrival time gaps are common for long-tailed distributions.

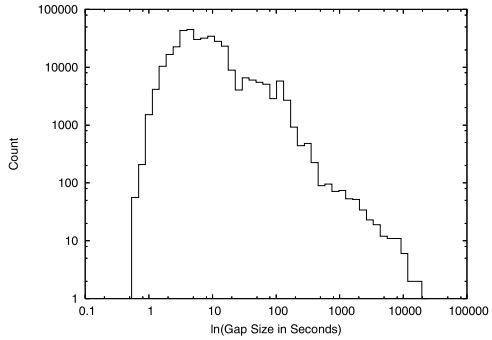
6.2. Network Churn

Network churn in the rate at which unique IP/port pairs appear or disappear from the network. The churn rate determines how fast a network map obtained using active probing falls out of date and becomes inaccurate. Churn was measured using both the active scan data and the passive mapping data independently because ground truth was not available. These analyses provide a conservative estimate of churn magnitude.

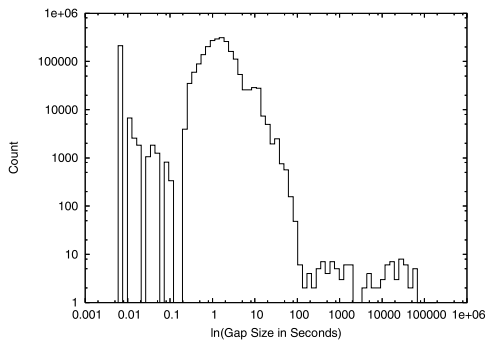
Active scan results were used to estimate churn by comparing results of the first and second scans. First, the data was checked to make sure that no subnets were seen by only one of the scans, as this could indicate that the rules on a subnet firewall were updated to either allow or block the scanning machine. Because no evidence of firewall rule changes was seen, all discrepancies between the first and second scan were assumed to be the result of actual network changes. Table 1 gives the results of comparing the services found in the first and second active scans. Services that are exclusive to the first scan (i.e. they were not found by the second scan) are assumed to have been removed from the network sometime between the two scans. Likewise, services exclusive to the second scan were assumed to have been added sometime between the two scans. In total, 93 changes occurred between the first and second scan. This is a rate of 1.2 changes per day.

Churn was also estimated using passive mapping data. Service connection counts were used to determine when each service first appeared on the network and when it was last used. One complexity of this analysis is that, when a service is first seen by the passive system, it must be counted either as a truly new, just initiated service or as a service

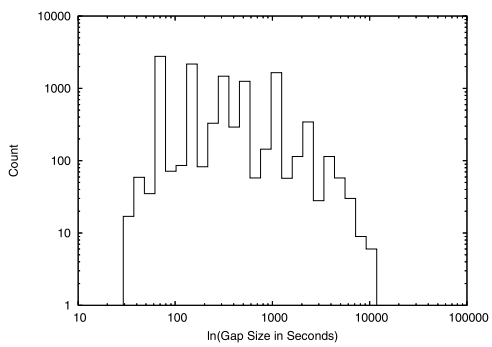
Fifth IEEE International Symposium on Network Computing and Applications (NCA'06)
0-7695-2640-3/06/$20.00 © 2006 IEEE
Authorized licensed use limited to: Klarquist Sparkman LLP. Downloaded on March 31,2026 at 20:01:57 UTC from IEEE Xplore. Restrictions apply.



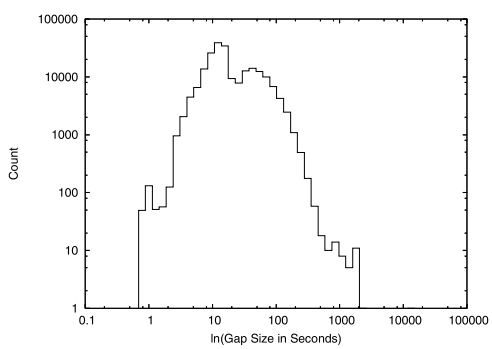
(a) HTTP: TCP Port 80 (342013 connection)



(b) VPN: UDP Port 10000 (2351423 connections)



(c) NTP: UDP Port 123 (11278 connections)



(d) SMTP: TCP Port 25 (208972 connections)

Figure 3. Histograms of connection inter-arrival time

Table 1. Network churn in active scan data

Scan	Services Found	Percent of Total
Total for First Scan	276	–
Exclusive to First Scan	32	9.5%
Total for Last Scan	305	–
Exclusive to Last Scan	61	18.1%
Total Changes	93	

that has been on the network since the start of data collection. This decision is made using the time from the start of data collection until the service is discovered and the distribution of inter-arrival times for that service. As discussed in Section 6.1, a threshold of $4 \times maxgap$, where $maxgap$ is the maximum inter-arrival time for that service, was used to differentiating previously terminated and missing services from services that started before passive analysis was begun. The following three criteria were required for a service to be considered new: (1) Service must have at least 50 connections (so that there are sufficient data points to accurately characterize the inter–arrival time), (2) The service was not seen in the first $4 \times maxgap$ seconds of data, (3) The service was not found by the first active scan. The requirements for determining when a service is no longer on the network were similar: (1) Service must have at least 50 connections, (2) The service was not seen in the last $4 \times maxgap$ seconds of data, (3) The service was not found by the last active scan. When evaluating the second requirement, the time where data is missing near the end of the data collection was not counted as part of the required "seconds of data".

Using the above conservative criteria, 57 changes were observed during the 76 days of monitoring (26 new and 31 terminated services). A total of 20 services were created and then removed while data was collected. Figure 4 shows a graph of changes over time. The relatively large discontinuities in the graphs near 59th and 72nd day are the result of the gap in the data. The "Service Added" line jumps just after the gap because all the services added during the gap were seen by the system for the first time once it started again. Similarly, the jump in "Services Removed" just before the gap is because, for all services removed during the gap, the last communication reported by the system was just before it stopped collecting data.

The temporal resolution of the passive system makes it possible to determine if terminated and added services are related. Mechanisms such as RPC, which dynamically assign services to ports, can cause services to change ports when a machine is rebooted. Hosts which change IP ad-

Fifth IEEE International Symposium on Network Computing and Applications (NCA'06)
0-7695-2640-3/06/$20.00 © 2006 IEEE
Authorized licensed use limited to: Klarquist Sparkman LLP. Downloaded on March 31,2026 at 20:01:57 UTC from IEEE Xplore. Restrictions apply.

dress, such as through DHCP, can give the appearance of services disappearing and then reappearing at a new IP address. Of the 57 total changes discovered with passive analysis, 30 appeared to be related in that one port disappears from a host at the same time as a new port appears. We did not find any evidence of hosts changing IP address because IP addresses are statically assigned in this DMZ network.

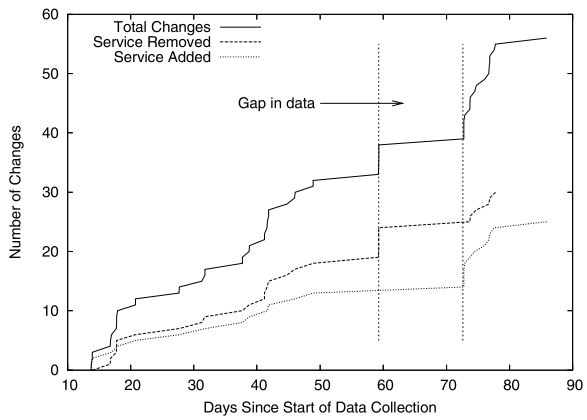


Figure 4. Cumulative changes over time

Active scan results show the network changing at a rate of at least once a day over the three month study. More conservative passive analysis estimates indicate a lower rate of a change every 1.5 days. These estimates are for a network with static IP addresses. Churn would presumably be greater for networks that use DHCP where IP addresses assigned to hosts change periodically, for networks that are changing in size, and for networks where operating systems are upgraded or changed frequently.

6.3. Service Discovery

Active and passive network mapping techniques differed in their ability to find services. As mentioned in Section 6.1, partial truth used to compare these approaches is derived by using services found by one system to evaluate the other system. To determine the performance of the active scanner, a set of services was chosen that, based on the data from the passive mapping system, were on the network at the time the active scan was performed. As discussed in Section 6.2, a conservative approach can be used with passive analysis to determine when a service is actually on the network. Because the first active scan occurred ten days before the start of passive data collection, there is no way to tell if a service found by the passive system was added to the network before or after the scan took place. Therefore, only the results of the last active scan were used for this analysis. The following two criteria were used to identify services that were present during the final active scan using data from passive analysis: (1) A connection to the service was seen after the scan finished, and (2) A connection to the service was seen during the week before the scan started. Services that match the above criteria were assumed to be on the network during the active scan. When looking for services missed by passive analysis, it was assumed that services found in both the first and last active scan were present over the time interval passive analysis was performed.

Table 2. All services found by one method only

Method	Total	Found	Missed
Passive	244	52 (21.3%)	192 (78.1%)
Active	105	43 (41.0%)	62 (59.0%)

Table 3. Eligible services found by one method only

Method	Total	Found By This Method Only	
		Number	% of Total
Passive	228	157	69%
Active	337	266	79%

Table 3 shows the results of the above analyses. In this table, "Total" is the total number of services in the baseline, "Found" is the number of services found by the listed technique, and "Missed" is the number of services not found. For example, the first row shows that 244 service/host pairs were in the first and last active scans but the passive mapping system discovered only 52 of these. This table shows that services discovered by one technique are largely missed by the other technique. Passive analysis did not find many of the services found by active scanning, but instead found a number of services that active scanning missed. An analysis of all discovered services presented in Table 2 supports this finding. This table shows the degree of overlap between all services found by the passive mapping system and all services found by active scans. In the table, the "Total" column shows the total number of services found by one technique and the "Number" and "Percent" columns under "Found By This Method Only" show the number not also found by the other technique.

Results shown in Table 3 and Table 2 demonstrate that service/host pairs found by active scans and passive mapping have little overlap on this network. This contradicts the assumption that passive traffic analysis could be used to update active scan results. An analysis of services missed by passive traffic analysis indicated that they included remote management ports on various pieces of network infrastructure blocked by the external firewall, default Windows services categorically blocked by the border router, or

Fifth IEEE International Symposium on Network Computing and Applications (NCA'06)
0-7695-2640-3/06/$20.00 © 2006 IEEE
Authorized licensed use limited to: Klarquist Sparkman LLP. Downloaded on March 31,2026 at 20:01:57 UTC from IEEE Xplore. Restrictions apply.

workstations that were officially used as clients but that had open ports that where used internally.

Generally, services were missed by the active scans because the machine performing the scans was blocked by a firewall. Some of the reasons why scans were blocked are:

- Firewalls protecting DMZ subnets often block connections from all other DMZ hosts.
- Firewalls protecting one or more special purpose machines block communication with those machines from everywhere except for a select set of remote hosts.
- Firewalls sometimes specifically block the central administrators scanning machine's IP address because local administrators don't want their network scanned.

6.4. Time To Discover Services Passively

A critical aspect of passive mapping is the time required to discover services. As discussed above, traffic rates and inter-arrival time distributions strongly affect this metric. The baseline for this analysis consisted of 59 services that were found by the first active scan and that were also discovered by passive analysis. This eliminates many services, but selects services that were almost certainly present from the beginning of the passive monitoring. Figure 5 shows a cumulative graph of the percentage of these services found during the 76 days of monitoring. The gap in monitoring, from day 49 to 62 on the plot, affects the curve only after day 49. This graph shows that roughly 50% of the services are found during the first day and 80% are discovered within the first ten days. The rate of discovery then drops off sharply with some services remaining undetected for more than a month. Figure 6 shows a log-log plot of the time it takes a service to be discovered versus the number of connections per day to that service. The roughly linear nature of the data indicate that the time to discover a service varies inversely with the connection rate. This graph shows that services that are connected to 1000 times per day are normally discovered in minutes, services that are connected to 10 times a day are normally discovered in hours to days, and services that are connected to once per day are normally discovered in days to weeks. Furthermore, the services that took the longest to discover are commonly used services, not custom applications. Services span a wide ranges of rates and it takes many days before passive traffic analysis discovers services with low traffic rates.

7. Discussion and Conclusions

This paper presents the first comparisons we are aware of between active and passive approaches to network mapping on an operational network. The above results demonstrate the capabilities of passive mapping, but they also raise important questions concerning the suggestion that passive mapping should be used to update active scans and to maintain an up-to-date map for network security situation awareness as suggested in [6, 7]. Hosts discovered by passive and active scanning in our study were mostly disjoint. In this environment, passive scanning is an important part of network situational awareness because it provides completely new information about services not seen by the active scans, not because it provides more timely information about services the active scans have already discovered.

Passive mapping also sometimes doesn't discover new services for hours or days. Passive mapping failed to detect more than 10% of known active services even after one month. In addition, the long tails for traffic inter-arrival times suggests that there is a finite probability of missing new services even when they have high average traffic rates.

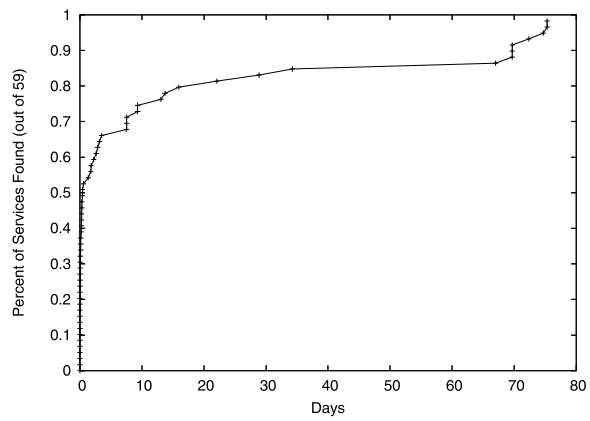


Figure 5. Time to find services

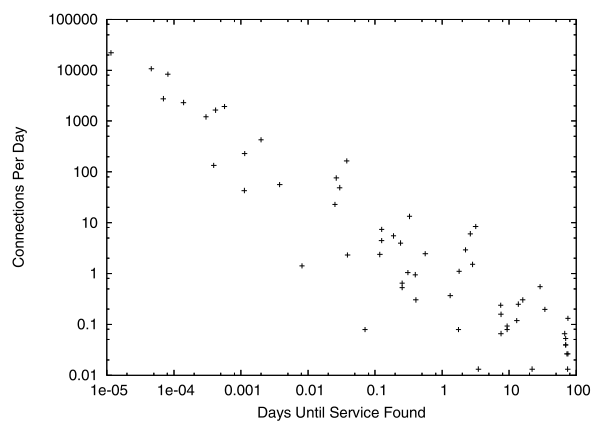


Figure 6. Time to discovery vs. number of connections

Fifth IEEE International Symposium on Network Computing and Applications (NCA'06)
0-7695-2640-3/06/$20.00 © 2006 IEEE
Authorized licensed use limited to: Klarquist Sparkman LLP. Downloaded on March 31,2026 at 20:01:57 UTC from IEEE Xplore. Restrictions apply.

These results demonstrate that it is not possible to rely either on normal connections to servers or on background scans and probes from the internet to expose services in a timely fashion for passive analysis.

Our analysis also examined the rate of churn and concluded that a service is added or removed at least once every one to two days. Anecdotal evidence from other enterprise networks suggests that these networks can have many more changes per day due to their size and complexity. With the high rate of malicious traffic on the Internet [16], even a single, new, vulnerable, public facing service may be discovered and exploited relatively quickly. These results suggest that active vulnerability scans may need to be performed daily to keep up with network changes, even when combined with passive mapping.

Our results suggest that relying on passive and active mapping together is better than relying on only active mapping alone, but that serious security risks remain. In particular, it is likely that there are services on the network studied not found by either active scans or passive mapping. Furthermore, services visible only to passive mapping may be discovered too late to prevent compromises. Even when hosts visible to active and passive mappers are identical, passive mapping does not provide instantaneous notification of new services, but can take a month to discover new services.

As noted above, our analyses provide an snapshot in time of one modern enterprise network. Further research is required to explore these analyses on other networks. It would be useful to determine whether active scanning is as restricted on other large networks and whether churn is higher on rapidly growing networks, networks with wireless components, and networks that use DHCP. It would also be useful to create inter-arrival time distributions for other services not seen on this network such as streaming audio and video, peer-to-peer networks, and chat clients. Finally, this work shows that, in some environments, there is no complete solution for defensively monitoring a network. Further research on scanning techniques that provide better network coverage without excessive instrumentation is required.

References

[1] Fydor, "The Art of Port Scanning," *Phrack Magazine*, vol. 7, Sept. 1997.

[2] "Nessus Security Scanner." http://www.nessus.org.

[3] The Subterrain Security Group, "Siphon Project," 2000. http://siphon.datanerds.net.

[4] M. Zalewski, "the new p0f: 2.0.5," 2004. http://icamtuf.coredump.cs/p0f.shtml.

[5] A. Ornaghi and M. Valleri, "Ettercap NG," 2004. http://ettercap.sourceforge.net.

[6] R. Gula, R. Deraison, and T. Hayton, "Passive Vulnerability Scanning," tech. rep., Tenable Network Security, Aug. 2003. http://www.tenablesecurity.com/images/pdfs/-passive_scanning_tenable.pdf.

[7] Sourcefire, "RNA Sensor," 2004. http://www.sourcefire.com/products/rna.html.

[8] V. Paxson, "Empirically-Derived Analytic Models of Wide Area TCP Connections," *IEEE/ACM Transactions on Networking*, vol. 2, no. 4, pp. 216–336, 1994.

[9] S. Sen and J. Wang, "Analyzing Peer-to-Peer Traffic Across Large Networks," *IEEE/ACM Transactions on Networking*, vol. 12, no. 2, pp. 219–232, 2004.

[10] L. Deri and S. Suin, "Ntop: Beyond Ping and Traceroute," *Proceedings of DSOM '99*, Oct. 1999.

[11] B. Dayioglu and A. Özgit, "Use of Passive Network Mapping to Enhance Signature Quality of Misuse Network Intrusion Detection Systems," *16th International Symposium on Computer and Information Sciences*, 2001. http://www.dayioglu.net/publications/iscis2001.pdf.

[12] R. Lippmann, D. Fried, K. Piwowarski, and W. Streilein, "Passive Operating System Identification From TCP/IP Packet Headers," *Workshop on Data Mining for Computer Security*, 2004.

[13] J. Kuntzelman, "Comparative Analysis of Active and Passive Mapping Techniques In an Internet-based Local Area Network," Master's thesis, Air Force Institute of Technology, Sept. 2004.

[14] A. de Montigny-Leboeuf and F. Massicotte, "Passive Network Discovery for Real Time Situation Awareness," *Proceedings of the RTO IST Symposium on Adaptive Defence in Unclassified Networks*, 2004.

[15] M. Crovella and A. Bestavros, "Self-Similarity in World Wide Web Traffic Evidence and Possible Causes," *IEEE/ACM Transactions on Networking*, 1996.

[16] The SANS Institute, "Survival Time History." http://isc.sans.org/survivalhistory.php.

Fifth IEEE International Symposium on Network Computing and Applications (NCA'06)
0-7695-2640-3/06/$20.00 © 2006 IEEE

Authorized licensed use limited to: Klarquist Sparkman LLP. Downloaded on March 31,2026 at 20:01:57 UTC from IEEE Xplore. Restrictions apply.