

A Discussion of Bandwidth Broker Requirements for Internet2 Qbone Deployment

Version 0.7

Editors:

Rob Neilson, BCIT

Jeff Wheeler, Nortel

Francis Reichmeyer, IPHighway

Susan Hares, Merit

Abstract

The purpose of this document is to present the requirements for a Bandwidth Broker, as defined within the context of the Internet2 Qbone Bandwidth Broker Advisory Council (BBAC). The requirements defined here are meant to provide guidelines for implementation of BBs to provide the QoS specified in [QBONEARCH]. In addition to the BB requirements, this document also proposes various deployment phases designed to provide a staged approach to providing the inter-domain communication necessary to achieve end-to-end QoS in the Qbone, using the mechanisms specified by the differentiated services WG in the IETF.

Both inter-domain and intra-domain issues are discussed, however, the primary focus is on the inter-domain requirements since these are recognized as the most important to being able to provide interoperation between BB on the Qbone. Many of the ideas in this document have been discussed previously in other documents, in particular BCIT's "Bandwidth Broker System Specification" [BCITBB], the IETF "two-tier" draft (draft-rotzy-two-tier-00.txt) [TWOTIER], and the papers from Olov Schelen, et. al. [AGENTS]. This document also intends to capture discussions on the i2-qbone-bb mail list and at the BBAC meetings.

THIS DOCUMENT IS A WORK IN PROGRESS.

A Discussion of Bandwidth Broker Requirements for Internet2 Qbone Deployment	1
<i>Abstract</i>	1
1. Terminology.....	3
2. Introduction.....	5
2.1 What is a Bandwidth Broker?.....	6
2.2 SLAs, SLSSs, and RARs.....	7
3. Deployment.....	8
3.1 Phase 0 - "Local Admission"	8
3.1.1 "Nailed Up" SLSSs.....	8
3.1.2 Two Service Levels	10
3.1.3 Static Inter-Domain Provisioning.....	10
3.1.4 Simple Intra-Domain Management	10
3.2 Phase 1 - "Informed Admission"	11
3.2.1 Static SLS Provisioning.....	12
3.2.2 Two Service Levels	12
3.2.3 Inter-Domain Communication	12
3.2.3.1 End-to-End Signaling	13
3.2.3.2 Immediate Response Signaling	13
3.2.3.3 Support For Both Signaling Models	14
3.3 Phase 2 - "Guaranteed Admission"	17
3.4 Phase N - "Dynamic SLS Admission"	18
4. BB Architecture.....	19
5. BB Requirements.....	21
5.1 User/Application Interface.....	21
5.2 Inter-domain Interface/Protocol	21
5.2.1 Operational Overview	22
5.2.1.1 Local Configuration.....	22
5.2.1.2 Session Initialization.....	23
5.2.1.3 Service Management	24
5.2.1.4 Resource Allocation.....	26
5.2.1.5 Session Maintenance.....	29
5.2.2 Minimum Performance Capabilities.....	30
5.2.3 Future Enhancements	30
5.3 Intra-domain Interface/Protocol	30
5.3.1 Interface Mechanisms.....	30
5.3.2 Intra-domain Interface Capabilities	30
5.4 Routing (Table) Interface	31
5.5 Multicast.....	31
6. References.....	31

1. Terminology

The following terms are used throughout the document and are defined here for clarity. Some of the terms defined in the DiffServ (DS) architecture [RFC2475] are repeated here for completeness. We also introduce some "new" terminology which is referenced throughout the document.

1.1 Diff Serv Terms

The following terms are found in [rfc2475]:

Downstream DS domain

The DS domain downstream of traffic flow on a boundary link.

DS boundary node

A DS node that connects one DS domain to a node in another DS domain or in a domain that is not DS-capable.

DS domain

A DS-capable domain; a contiguous set of nodes which operate with a common set of service provisioning policies and PHB definitions.

DS egress node

A DS boundary node in its role of handling traffic as it leaves a DS domain.

DS ingress node

A DS boundary node in its role of handling traffic as it enters a DS domain.

Service

The overall treatment of a defined subset of a customer's traffic within a DS domain.

Service Level Agreement (SLA)

A service contract between a customer and a provider that specifies the forwarding service a customer should receive. A customer may be a user organization (source domain) or another DS domain (upstream domain).

Service Provisioning Policy

A policy that defines how traffic conditioners are configured on DS boundary nodes and how traffic streams are mapped to DS behavior aggregates to achieve a range of services.

Upstream DS domain

The DS domain upstream of traffic flow on a boundary link.

1.2 "New" Terms

In addition to the terms from [RFC2475], we define the following:

Bandwidth Broker (BB)

A bandwidth broker (BB) manages network resources for IP QoS services supported in the network and used by customers of the network services. A BB may be considered a type of policy manager (see Policy Manager definition below) in that it performs a subset of policy management functionality. (edit done Rudiger)

Connection Admission Control (CAC)

Connection admission control refers to the process, performed by the BB, of admitting connection requests to the network based on available resources in the network. The determination of available resources may be done on a static or dynamic basis.

Domain

A domain typically refers to DiffServ domain - see *DS domain* above, from [RFC2475].

Edge Router (or Edge Device)

We use the terms edge router, edge device, and boundary node interchangeably. See *DS boundary node* above, from [RFC2475].

Inter-Domain Communication

Inter-domain communication refers to the protocol messages and control data that are exchanged between BBs in adjacent domains. (edit done Ben)

Intra-Domain Communication

Intra-domain communication refers to the protocol messages and control data that are exchanged between a BB and the nodes (usually edge devices) within that BB's domain. (edit done Ben)

Peer Domains

Two domains are peer domains if they are adjacently connected.

Per Hop Behavior (PHB)

The externally observable forwarding behavior applied at a DS-compliant node to a DS behavior aggregate. Note that while each service is mapped to a PHB (and specific DS Code Point(s)), it is not possible to identify a service by it's PHB (e.g. AF).

Policy Manager (PM) or Policy Server (PS)

A policy manager (PM) or policy server (PS) typically manages the access of users to network policy services. As part of the process of admitting users to access policy services, a PM may employ a BB for CAC, as described above.

Premium Service

Premium Service refers to a quantitative differentiated service which provides a guaranteed low loss and jitter over a DS region. The Premium Service often is also described as "Virtual Leased Line (VLL)" Service. The exact service specification may be found in [QBONEARCH].

Resource Allocation Request (RAR)

A RAR refers to a request for network resources (or service) from an individual user to the BB of that user's domain. If the request includes network resources for t outside of the user's local domain, the admission control may be performed based on the SLS(s) in place with adjacent domains. Accepted RARs may result in service provisioning policy (see above) installed in edge devices by a BB. (edit done Ben)

Service

Service is the overall treatment of a defined subset of a customer's traffic within a DS domain or end-to-end [RFC2475]. In this document, the [RFC2475] service definition will also be applied for traffic treatment between two domains. This leads to unilateral, bilateral and end-to-end service specifications. Whenever "service" is used as stand alone term in the following, bilateral and end-to-end services are meant. Each "service" is mapped to a PHB identified by its DS Code-Point(s) (DSCPs). By this definition a "SERVICE" IS IDENTIFIED BY ITS "DSCPs" within a DS domain as well as between two adjacent DS domains in the following. The IETF does only standardize PHB's. IETF specifications usually DO NOT LINK DSCPs TO SPECIFIC SERVICES. While each service is mapped to a PHB (and specific DS Code Point(s)), it is not possible to identify a service by it's PHB (e.g. AF). (edit done Ben and Ruediger's)

Unilateral Service

Unilateral service is used to refer to "service" as defined in DiffServ [RFC2475] (above).

Service Level Agreement (SLA)

See SLA in [rfc2475] which defines SLA as " a service contract between a customer and a service provider that specifies the forwarding service a customer should receive. A customer may be a user organization (source domain) or another DS domain (upstream domain). A SLA may include traffic conditioning rules which constitute a TCA in whole or in part.

Service Level Specification (SLS)

An SLS refers to the particular information relative to the BB and the network devices in order to support a SLA in that network. Information in an SLS is generally on the level of aggregate data flows and the resources/bandwidth provisioned for those flows. An SLS is typically applied at the endpoints of a link connecting adjacent domains and reflects traffic that will be sent from the upstream domain to the downstream domain.

Service Users

End systems users and other entities that can generate RARs. It could as well be an operator that does the RARs (e.g. after being contacted by end-users).

Subnet Bandwidth Manager (SBM)

A Subnet Bandwidth Manger is in charge of the resource allocation requests for a subnet. All users on a variety of hosts on a subnet would defer to the Subnet Bandwidth Manager to negotiate the bandwidth with the Bandwidth Broker within a domain. The communication path to request resources would be the host signaling the SBM that it needs premium service. The SBM will send an RAR to the BB within the domain. An SBM can also be pre-configured with the ability to requests certain bandwidth resources.

Virtual Leased Line (VLL)

See Premium Service.

2. Introduction

This document presents some high level specifications for bandwidth broker functionality, as part of the efforts of the Internet2 Qbone Bandwidth Broker Advisory Council to specify and define what a bandwidth broker is and how it will be deployed in the I2 Qbone test network. The specifications, as presented here, are meant to initiate the exchange of ideas, among the Advisory Council, on bandwidth brokers and to establish some agreements on bandwidth broker issues for the purpose of preliminary implementation and interoperability testing.

In [QBONEARCH] the term bandwidth broker (BB) is used to refer to the abstraction that automates the admissions control decisions and network device configuration functionality according to a common set of operational practices and procedures specified of the network, in the context of the Qbone architecture. In this document, we expand upon that concept and discuss the requirements of a bandwidth broker for the purpose of inter-domain communication to achieve the end-to-end Qbone services discussed in [QBONEARCH].

Readers familiar with the ongoing efforts within various IETF policy- and QoS-related working groups (e.g. Policy Framework, RAP, Diff Serv, and AAA) may notice some similarities between the functionality of the bandwidth broker (BB), described here, and that of a Policy Server (or PDP [RAPFRAME], or Policy Services Manager, or some other similar name) as is being defined within the IETF. Connection Admission Control is recognized as a necessary part of providing QoS and policy services. Ideally, bandwidth brokers and policy managers will work together to provide call admission control in an integrated policy services environment. Still, there are several functions that a policy manager provides (or may provide) that are not necessarily required of a bandwidth broker but that a BB may utilize via an

interface to a policy manager [Indeed, a BB may even be built into a policy manager or considered a type of policy server]. Examples of this include complex policy conflict resolution and SLA management (where SLA is used in the sense of a contract between provider and customer).

In the absence of an interface with a policy manager, some basic policy or service level specification information is needed by the BB. This is included in this document as BB specifications with the assumption that policy managers are not considered part of the network architecture at this time. However, it is *not* the intent of this document to recommend policy manager requirements, in general.

What was just about a policy server in principle also holds for a network management system. The BB may be built into the network management system or be considered as a type of network management system. In the absence of an interface with a network management system, some basic network management or service specific resources (and may be performance) information is needed by the BB. This is included in this document as BB specifications with the assumption that network management systems are not considered part of the network architecture at this time. However, it is not the intent of this document to recommend network management system requirements, in general. (Ruediger's comment)

While clearly beyond the scope of this document, it should be noted that also policy server, BB and network management system may be integrated into a single system. A separation of the tasks policy-, service- and network management may be require interfaces and protocols allowing a shared access to such a system. (Ruediger's comment)

2.1 What is a Bandwidth Broker?

We've mentioned that a BB can be considered a type of policy server, and we've also mentioned that a BB is not necessarily a policy manager but policy management and bandwidth brokering will need to work together in providing integrated policy services and call admission control. So, what is a bandwidth broker and what does it (or should it) do? This section addresses that question.

In short, a BB manages the QoS resources within a given domain based on the Service Level Specifications (SLS) which have been agreed in that domain. The SLS is a translation of a Service Level Agreement (SLA) into the appropriate information necessary for provisioning and allocating QoS resources within the network devices, and in particular at the edges of the domain on links connecting the domain to adjacent domains. The BB is also responsible for managing inter-domain communication, with BBs in neighboring networks, for the purpose of coordinating SLSs across the domain boundaries.

The BB gathers and monitors the state of QoS resources within its domain and on the edges of the domain going to and from adjacent domains. That information, together with the policy information (from the policy rules data base) is used for admission control decisions on QoS service requests to the network. The policy manager, if it exists, verifies such requests against installed policy rules, checking for conflicts with existing requests, taking appropriate preemption actions, etc. The network state information from the BB is also used to verify that resources are currently available in the network to support the request. This information may be obtained directly from the BB via an interface to the policy manager or, more likely, the BB places it in a common database shared by the policy manager.

Across boundaries, the SLS may be on an aggregate basis, where aggregation is on all flows within the domain of a particular QoS service type (i.e. DiffServ codepoint). Within a domain, individual flows may be allocated resources from the SLS by issuing Resource Allocation Requests (RARs). It is the responsibility of the BB to coordinate allocation and provisioning of the aggregate resources of the SLSs, into and out of its domain, with those resources requested from RARs.

Figure 1 shows a sample network configuration. It consists of three domains AS1, AS2, and AS3 with bandwidth brokers BB1, BB2 and BB3, respectively. Also shown are the SLSs in place between AS1 and AS2, and between AS2 and AS3 as well as individual service users attached to AS1 sending RARs to BB1.

In this case, a service users is an end system or application which requests bandwidth. (see the definition of service user in section 1). This sample configuration will be referred to as we describe various possible stages of implementation in the next section on deployment of QoS and BBs in the Qbone.

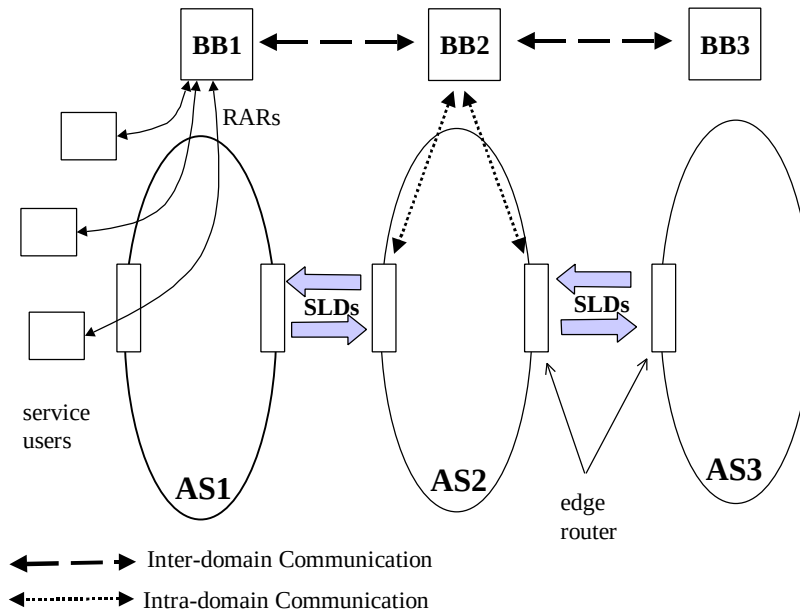


Figure 1: sample network configuration

2.2 SLAs, SLSs, and RARs

It is important to understand the relationship between service level agreements, service level specifications, and resource allocation requests, as mentioned above. In general a SLA is a contract that specifies a forwarding service a customer can expect to receive from a provider. A particular SLA may cover one or more services the provider offers to the customer and part of the agreement will specify how the provider is to recognize, or classify, packets from the customer which are to receive which service. For example, in the case where the customer and provider are peer provider domains, as opposed to individual users and a provider, the SLA could specify different levels of service(s) to be provided on traffic aggregates. Each traffic aggregate could be designated by particular DSCPs in the packets marked by the customer before sending the packets to the provider.

Service Level Specifications (SLS) describe the resources provisioned for a particular service, specified in a SLA, across a domain boundary link. A SLS will also specify the classification for packets that are to be (re)marked to receive the particular service in the downstream provider network, for example by DSCP. Optionally, a SLS may also specify the shaping characteristics to be applied to the traffic as it leaves the domain. One SLS is specified for one particular service; if a SLA is written that covers more than one service offering then multiple SLSs are generated for that SLA - one for each service and link between domains (in the case of more than one link between two domains).

Resource allocation requests (RAR) are sent to the BB by service users to request the use of the resources provisioned within the domain for a particular service. The RAR contains the destination of the traffic for which the request is made and, from that, the BB can determine whether the traffic is local (the destination is also in the BB's domain) or transient (the destination is in another domain). If transient, the downstream domain is obtained and the resources specified by the SLS on the egress router are allocated for that request. Several RARs may be mapped to a single SLS when the bandwidth of RAR requests fits within a

single SLS. When the resources provisioned in a SLS are fully allocated to existing RARs and a new RAR is received for the same destination, the RAR must be rejected or the SLS must be re-negotiated (either statically or dynamically; see Section 3 below).

3. Deployment

This section describes some ideas and recommendations for BB deployment in the Qbone. "Phase 0" refers to initial deployment intended to start out and gain interoperability experience, etc. "Phase 1" refers to the next steps that are likely to be taken, etc. The main focus for these deployment phases is on the inter-domain behavior since it is necessary to have those requirements agreed upon for interoperability and testing in the Qbone. For each phase, the intra-domain behavior is mentioned but it is less important, at this time, that all BBs support the same functionality within their individual domains, as long as the inter-domain requirements are met.

[Editor's note: Please review sections 3.1 for based on the notes from the 6/24/99 phone conference]

3.1 *Phase 0 - "Local Admission"*

Phase 0 is referred to as "local admission" because from the BB point of view, requests for resources are admitted into the network with admission control relative only to the local domain where the request is received. If the request is accepted in the local domain it is admitted into the network without further checking if resources will be available all the way to the actual destination. The highlights of this phase of deployment are listed here and described below.

- "nailed up" SLSs, network engineered to support SLSs by configured necessary routers with static inter-domain provisioning, e.g. "70/30 rule"
- two service levels: "Premium" and best effort
- simple intra-domain management, e.g. "count to ten"
- intra-domain BBs protocol is a local issue; BB configures routers within it's domain
- Host BB Interaction

3.1.1 "Nailed Up" SLSs

Initially, SLSs will be established statically and edge routers in peer domains that share an SLS will be configured to service packets (police, mark, queue, schedule, shape, etc.) according to the DiffServ information in the SLS. The amount of resources/bandwidth specified in an SLS must be agreed upon by both sides. The resources for the SLS may be provisioned, for example, as some percentage (e.g. 30%) of the total capacity of the link connecting the two routers. There is no inter-domain communication between the BBs yet, so the SLS negotiation will be performed manually (telephone, email, etc.) by the network operators of the two domains. The SLS will then be installed in the BB via the network operator's GUI into the BB and/or into the affected edge routers.

Figure 2, provides an example of the SLSs set-up between two end-sites requesting Premium service. User 1 in domain 1 asks for Premium service to User 3 in domain 3. User 2 asks for Premium service to User 4 in domain 4. In domain 1, the BB1 will configure routers along the pathway to the edge router attached to domain 2 to support the requested reservation of resources. The BB to router communication may use COPs, or a Command line interface or SNMP to communicate the necessary information to configure the PHB.

The router configurations initiated by BB1 to support the service reservation request of User 1 may differ depending on the location of a router in domain 1: Router 1 may be configured to classify and mark packets received from User 1 with the appropriate DSCP. Router 1 further may be configured with the appropriate policing parameters and it may be configured to shape User 1's flow. Assuming pre-configured Premium resources within domain 1, router 2 may not have to be updated at all. Router 3 may need to be configured with different shaping parameters before it passes on User 1's traffic to domain 2. If and where a BB

configures DiffServ specific router parameters depends on a Internet-Providers service management policy as expressed in the SLA and SLDs.

In domain 2, by manual pre-configuration BB2 has a pre-configured SLD. BB2 configures Routers 4, 5 and 6 in domain 2 to allow a particular premium service code point. Like wise, BB 3 will configure routers 7 and 8. BB 4 will configure routers 9 and 10.

At the inter-domain boundaries the edge routers will have a manually configured SLS for the inter-domain Premium service bandwidth. In figure 2, router 3 and 4 are manually configured to have the appropriate SLS for the inter-domain Premium service bandwidth. In figure 2, router 3 and 4 are manually configured to have the appropriate SLS for the inter-domain Premium service bandwidth. In figure 2, router 3 and 4 are manually configured to have the appropriate SLS for the inter-domain Premium service bandwidth.

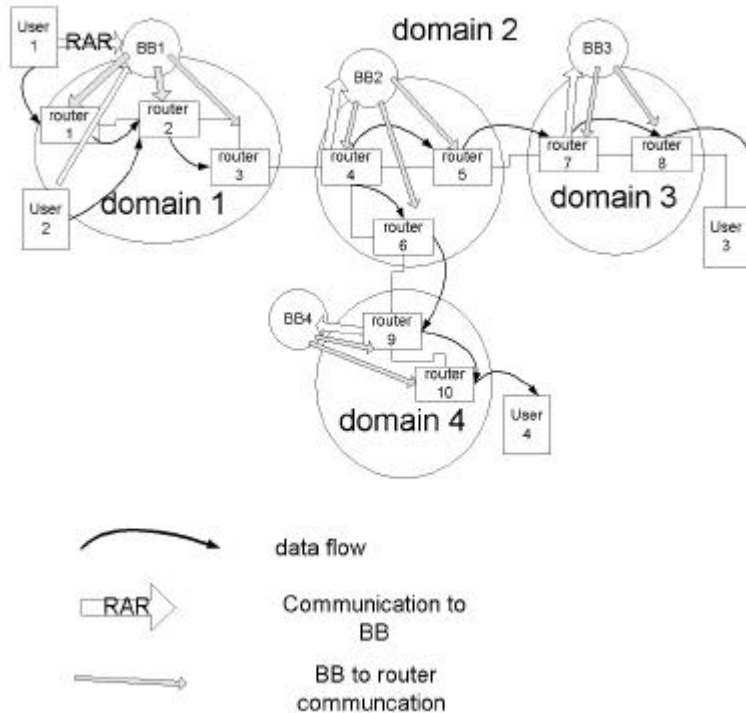


Figure 2 - Data flow and BB communication in Phase 0

There is no inter-domain communication between BBs yet, so this SLD negotiation will be performed manually (telephone, email, etc.) by the network operators of the two domains and will be installed in the BB via the network operator's GUI into the BB.

Routers within a domain may optionally indicate to BBs that a packet with a premium service (a particular DSCP) has arrived. This communication only provides information to the local management of the BB and does not directly influence the BB communication.

The communication of the RAR request for the User may alternatively come from a Subnet Bandwidth Manager (SBM). The User located on a certain subnet would signal the SBM to request bandwidth resources from the Broker. The SBM would request the resources from one of the domain's BBs. The BBs would then configure the routers just as in figure 2. Optionally, the SBM may be pre-configured to request bandwidth upon noticing that a particular host is up.

3.1.2 Two Service Levels

For Phase 0 deployment, only two service levels will be supported: Premium service and best-effort service. A specification of the Premium service, for this phase, is given in [QBONEARCH]. Within a

domain, originating hosts or edge routers mark packets with a DSCP associated with Premium service for that domain.

Premium service as specified by the Qbone Architecture document does not allow remarking of packets. However, in the process of initial start-up of the Qbone it may be necessary to remark packets to allow the PHBs to function as desired. For this case in across domain boundaries, the SLS may specify that the egress edge router of the upstream domain (re)mark packets with a DSCP associated with Premium service in the downstream domain. Alternatively, the SLS may specify that the ingress edge router in the downstream domain will classify packets with the upstream domain Premium service DSCP and do the remarking.

3.1.3 Simple Intra-Domain Management

The phase 0 intra-domain BB efforts will likely end up being simplified as well. Network resources may be represented in the BB as an amount of resources that can be allocated to all QoS flows entering the network. This amount would be chosen intelligently based on the network engineering used to provision and configure the network to provide the services in the first place. The total available specific resources in the network may be divided among ingress points as desired by the BB. For example, it could be shared equally among all ingress interfaces, or allocated on first-come-first-serve basis, etc.

As new RAR requests are accepted into the network and old ones are removed, the amount of resources available is adjusted accordingly. The BB performs CAC on these requests using whatever method is appropriate for determining available resources in the domain. If the CAC passes, the BB assumes the network has been properly engineered, internally and at the egress points, to service the request.

Complex policy management is not expected to be required in this beginning phase. This is because the number and types of services that will initially be offered will be small (Premium and best effort), as will the number of different customers, and the inter-domain requests are of a static nature. All of these factors mean that the policy associated with these requests will initially be manageable by a BB without the need for complicated conflict resolution or techniques for flow preemption based on policy.

3.1.4 Intra-domain BB protocol

In order to coordinate bandwidth assignment across domains, a single inter-domain BB protocol must exist.

Within a domain, the BB protocol or protocols must simply inter-operate within that domain. A single intra-domain BB protocol may be used to exchange information between a host and a BB, a Subnet Bandwidth Manager and a BB, two BB within a the same domain, and a BB and the routers providing the premium service. Or alternatively within a domain, a BB may be configured to use any set of protocols to communicate with the hosts, routers, SBMs or other BB. The BB will communicate with either the SBM or User via RAR to:

- receive the request for Bandwidth,
- process the bandwidth request with all other BBs within a domain,
- signal the appropriate inter-domain BB requests given the bandwidth requests,
- indicate success or failure to obtain bandwidth to the User or SBM.

Routers within a domain may optionally indicate to BBs that a packet with premium service with a particular DHCP has arrived.

Examples of some protocols which can be used to communicate with routers are : Diameter, COPS, a simple Command Line Interface (CLI) based. Current implementations of BBs use COPS (UCLA), CLI (Merit, BCIT), and SNMP (Telia) to change the routers to provide the premium service requested.

Examples of some protocols which can be used to communicate with hosts include: RSVP, COPS, a simple Command Line Interface (CLI), and Diameter. Intra-Domain BB protocols include: Diameter, COPS and others.

[note: Editorr: Do we have a complete list of the intra-domain BB protocols that we know about?]

3.1.5 Host to BB interaction

The Host may send RAR to the BB directly and receive back indications of bandwidth set-up. If the bandwidth resources is allocated, the Host can begin to send data. If the bandwidth resource is not allocated, the Host must indicate to the user or application that the bandwidth request is not being fulfilled. Optionally, on a failed request the BB may pass "hints" when the bandwidth might be granted.

The Host BB interaction may be through a SBM instead of directly. The host might just send packets and the SBM would capture the packets, and perform the necessary resource allocation. Another option is that the SBM might detect the Host on the network and start the resource allocation process based on some pre-configured information.

3.2 Phase 1 - "Informed Admission"

Phase 1 is referred to as "informed admission" because the local BB makes admission control decisions based on information obtained from downstream BBs. Depending on the service being requested, the information could be obtained from just the next hop BB or it could come from, in a sense, all BBs along the path to the destination. This information is obtained via an inter-domain signaling protocol used between adjacent BBs. **[Note: Details of the inter-domain protocol, message contents, etc. will be provided in a later section of this document.]** The highlights of this implementation phase are listed here and described below.

- Static SLS provisioning, as in Phase 0
- Two service levels, as in Phase 0
- Intra-domain BB is a local issue, as in Phase 0
- Inter-domain communication, notification/response protocol
 - provides service specific resource administration
 - allows for internal, manual provisioning and SLS adjustment
- Signaling information based on resource allocation and provisioning only
- AAA
- Security issues
-

3.2.1 Static SLS Provisioning

For Phase 1, SLSs are still "static" and manually coordinated with the engineering and configuration of the network. The inter-domain communication (Section 3.2.3) is not meant to provide dynamic renegotiating of SLSs across domain boundaries. Instead, the BB-to-BB signaling is intended to provide BBs with information regarding destination(s) of aggregate traffic being sent from one domain to the next, and available resources in the downstream domain(s). This information could be used, at this phase, to internally assess resource availability and possibly trigger a network management application if SLSs need to be re-negotiated.

The problem with requiring dynamic SLS re-negotiation by the BBs is that adjusting resource commitments at the edges (SLSs) requires reconfiguration of the internal network devices as well, for example to provision more resources for Premium service (see Section 3.4). By limiting the requirements for this phase to simple notification and response signaling, we provide the mechanisms to allow developers to experiment with different intra-domain brokering techniques while not requiring them.

3.2.2 Two Service Levels

Phase 1 deployment will support two service levels, Premium service and best-effort service, as in Phase 0 (see Section 3.1.2).

3.2.3 Inter-Domain Communication

Phase 1 deployment will support notification and response model for inter-domain communication. Two possible signaling models are shown in Figure 2: End-to-End signaling and Immediate Response signaling. BB1, BB2, and BB3 represent the BBs for AS1, AS2, and AS3 shown previously in Figure 1. A third alternative of a "Central Manager" has been considered but left for future development (see section 6 on Future developments).

Inter-Domain resource reservations are uni-directional from a source to a destination. Two applications programs desiring bi-directional resource reservation will need to create this from two uni-directional reservations. We anticipate that the application programming interfaces (APIs) will specify this functionality and the underlying application middleware will provide the combining of two reservations of the same type.

Editor's note: Ruediger Geib will provide additional text for this section based on VLL inside of other VLLs.

3.2.3.1 End-to-End Signaling

Figure 2A shows the end-to-end notification/response signaling. Here a user request (not shown) is received by BB1 in domain AS1. After determining that local resources are available and the request will not violate the SLS with the downstream domain (just as is done in Phase 0), BB1 notifies BB2 of the effect of the request on the aggregate resources provisioned in the SLS between the domains. For example:

1. BB1 notifies BB2 how much (aggregate) traffic will cross the boundary into AS2 as a result of accepting the request, and the destination AS(s) of the aggregate
2. BB2 does its local admission control and notifies BB3, the downstream AS - and in this example the destination AS.
3. BB3 verifies resources for the request in its local domain and sends a (positive) response to BB2.
4. BB2 forwards the response to BB1 which accepts the service user's request.

If local admission control fails at any along the way, notifications stop flowing downstream and a (negative) response containing the AS_ID where the failure occurred is immediately sent back upstream to the originating BB where it is processed and the appropriate action taken. In figure 2, this failed local admission is shown for Domain 2 in the response 2b. The initial request is sent from BB1 to BB2. If BB2 cannot provide the local admission, then response 2b is sent back to BB1.

3.2.3.2 Immediate Response Signaling

Figure 2B shows the immediate response model for inter-domain communication. In this model:

1. after performing local admission control as described above, BB1 send the notification to BB2 with the aggregate bandwidth usage, destination AS information, etc. and BB2 does its local admission control

2. If admission control fails at this point (e.g. not enough resources available in AS2), BB2 sends a negative response to BB1 and we're done. If BB2 "accepts" the notification, a positive response is sent back to BB1 (go to step 3). Upon receiving the response, BB1 passes it to the service user (action 2b).
3. BB2 sends notification to BB3 which performs local admission control
4. If BB3, being the destination AS, accepts the traffic from BB2 it simply responds to BB2.
5. If BB3, being the destination AS, denies the traffic from BB2 it responds to BB2 with a negative Acknowledge for allowing the traffic to enter BB3.
6. If BB2 receives a negative acknowledgement from BB3, it passes the negative acknowledgement to BB1.
7. BB1 passes the negative acknowledgement to the service user.

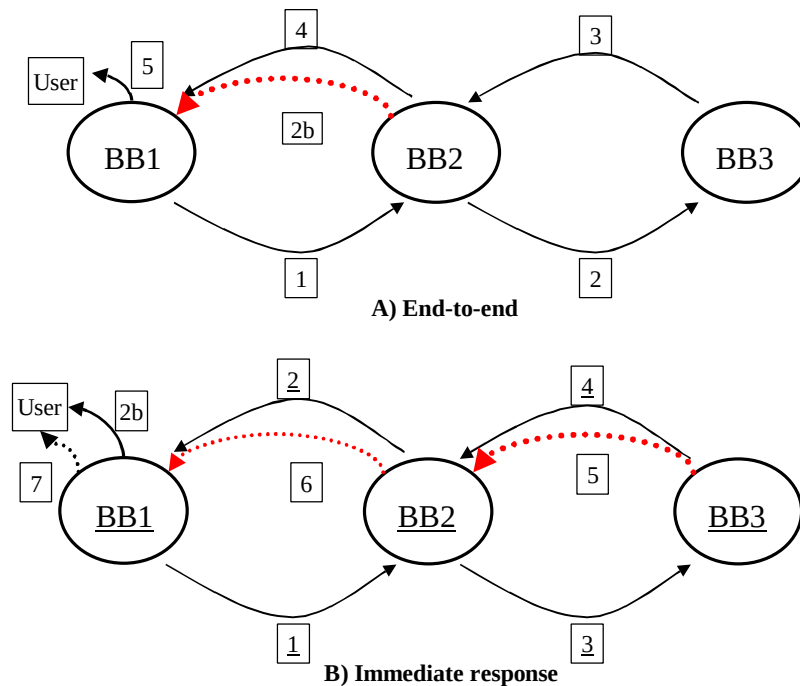


Figure 2: Inter-domain Notification/Response Examples

In the immediate response model, we see that BB1 may accept the request from its local user upon receiving a positive response from BB2 and without any indication that QoS resources for the traffic flow will be available in any other downstream ASs. Thus, traffic may be sent through AS1 and AS2 only to be dropped in AS3 due to insufficient provisioning in that network. This is not possible in the end-to-end model since notifications and responses go from BB-to-BB to the destination AS and back before the user's request is admitted. Which of these models (and there may be others, hybrids or variations) makes the most sense may depend on the service being requested or the application requesting the service. If a "guarantee" of end-to-end QoS is required before traffic is sent, the model in Figure 2A would be more appropriate. However, if an application simply wants to get the best service, as far as possible, the immediate response model of Figure 2B would suffice.

3.2.3.3 Support For Both Signaling Models

It may be desired, or necessary, to support both end-to-end and immediate response signaling models. One reason for supporting both signaling models is that different applications might require different levels of information about resource availability along the path from sender to receiver.

End-to-end signaling requires that the whole network path be set-up prior to sending any data out of the originating domain, that is the domain that sources the data to the QoS infrastructure. Immediate response allows data to flow out of the initial domain into an adjacent domain once the adjacent domain has acknowledge that it has resources to carry the traffic. If some portion of the complete service pathway does not have enough resources to carry the data to its destination, the data will be dropped and the complete pathway torn down.

End-to-end signaling provides added value if the probability of obtaining the full pathway resources is mixed and traffic has a fair to good chance of being dropped if immediate signalling was being used

Immediate signalings allows data tranfer to take place while the negotiation is occuring if the following assumptions are true:

- the probability of obtaining the full pathway resources is high (90% or greater is a good guess),
- the BBs can process the requests and reconfigure the routers before packest traverse the domain,
- the reservation system consistently performs in this manner within a domain.
-

and the BB system can process the request by configured before packets traverse the domain If any of these assumptions are not met and the per-hop-behaviors are not set-up in each router, packets will be dropped in the first router which is not configured along the data transfer pathway.

This document does not discuss inter-operability between the end-to-end signaling model and the immediate signaling model. We anticipate that a service will be built out of either an immediate response signaling or end-to-end signaling service. For example, some applications may need to know that resources for the requested service can be obtained all the way to the destination; sending application data that might get dropped or delayed in best-effort delivery due to insufficient resources is not acceptable. An application built on the Premium service to maximize the PHBs of bandwidth and delay may require that such a service be established on an End-to-End basis prior to sending data outside of the initial domain.

Some applications could make use of better-than-best-effort service as much as possible in the network and having packets dropped or down-graded to best-effort at points along the path will not compromise the application. For these applications, immediate response signaling is sufficient and the end-to-end model would be overkill. Applications making use of the AF service or applications running over an over-built infrastructure may desire the immediate response service.

[Editorial change: The packets will be dropped, not down graded. We need to discuss dropping this paragraph.]

To facilitate both possibilities, we suggest the use of a flag in the inter-domain protocol notification message to indicate the type of response method requested. In individual sessions between two bandwidth brokers, different SLS may request different types of signaling. A flag in the Session Service Request (SSR) and the Set-up Session Answer (SSA) indicates the mode, end-to-end signaling or immediate signaling that each session is engaging in. Please refer to section 5.2.1.3 for details.

3.2.4 Signaling Information based on Resource Provisioning and allocation only

In this phase, the first introduction of inter-domain communication, we consider the exchange of information relative to resource provisioning and allocation only. That is, this phase of signaling is for facilitating authorization of a request based on available network resources. Information for the purpose of authentication of end users/applications is not considered at this time. Any authentication of this sort is assumed to be handled intra-domain (authentication stops in the sender domain, being performed on the RAR), or to be handled separately via some other remote peer-to-peer signaling mechanisms. Of course, the signaling messages themselves should be authenticated so a BB can verify the identity of peer BBs to ensure they are allowed to make resource requests.

[editor: This next section has been suggested by Ruediger Geib, and will be discussed at the next meeting.]

End to end signaling based on a sequential and strictly bilateral mode of inter domain BB communication may require intermediate BB's to "route" also end to end signaling messages (no admission decision required or expected). This is required as the current inter BB protocol only supports communication between BB's of adjacent domains. To inform an intermediate BB about the "routing only" requirement, a special parameter or a special message type may be used.

Alternatively, a separate end to end communication may be installed. The end to end communication is surely not a BB phase 0 issue and it remains for future study.

3.2.5 AAA features

Bandwidth Brokers(BBs) manage a precious commodity in the Qbone network, the premium bandwidth. Early phases of the Qbone will use simple authentication mechanism to make sure the each BB peers with a valid peer. Some of these authentication methods could be as simple as the use of a password or MD5 encoding. Authorization and Accounting functions may be non-existent or very rudimentary during the early stages of the Qbone. Since the focus of the early Bandwidth Broker is a functioning bandwidth broker, the securing of common AAA functions within bandwidth brokers may take a back seat to the basic bandwidth broker functions.

As the Qbone, moves from to later phases of deployment it is anticipated that the administration of these bandwidth brokers will include stronger Authentication, Authorization and Accounting (AAA) functions. Just as the dial-in connections within the Internet progressed from simple dial-in with little or no administrative control to more extensive AAA functions, BBs will add AAA functions as they mature. The IETF is working on extending AAA functions to support Bandwidth Brokers. Many diff-serv domains might want to experiment with portions of the AAA protocols with a co-operating diff-serv domain. In order to enable these valid experiments, the inter-domain BB protocol will allow optional fields to support the functions within an inter-domain protocol. Experiments have been suggested by members of the I2 Qbone BB Advisory Council on authorization and accounting methods.

To summarize, the use of AAA functions within a Domain is up to the administrator of the domain. The domain may be physically secure or experimental in nature so that no AAA functions are required. The AAA functions can be non-existent, weak or strong.

Inter-Domain AAA functions are optional between peers. The only requirement for all BB peers is that you use general method of validating your peers. Peer validation may be based on pre-configuration or network management. Optionally, two diff-serv domains may co-operate in their BB exchange to experiment with stronger BB functions.

[This section has been discussed by Susan Hares and Rob Nielson. We will discuss this at the next BB meeting]

3.2.6 Security issues

The security aspects associated with inter-domain communications: securing the message and validating that you have a valid peer. Pre-configuration of a peer can be viewed as a form of a simple method of validating the (message and the peer). Strong communication requires peer-to-peer identity and phase of that, while are optional at the end of the signaling information may be secured, for example using IPsec: BB-to-BB (peer BBs) and end-to-end. In BB-to-BB security, the security association exists only between the two BBs involved; messages encrypted at the sending BB can be fully decrypted at the receiving BB and the receiving BB has access to all data in the message.

End-to-end message validation forms no part of this document yet. In an end to end security model, the (original) sending BB encrypts the message such that only the destination BB can decrypt and access the complete data in the message. An end to end signaling model may require intermediate BBs to process the message (e.g. for routing purposes or CAC).

Any security mechanism should support separated security mechanisms for specific contents of the inter-domain BB signaling parameters.

[editor - need comment on the interaction of this with AAA.]

End-to-end BB security , the (original) sending BB encrypts the message such that only the destination BB can decrypt and access the data in the message; intermediate BBs access the message headers for routing purposes only. The obvious problem with the end-to-end model is that BBs are managing resources on an aggregate basis and must request resources from downstream domains based on aggregates of all upstream requests received. Thus, data in upstream messages must be processed at each BB. Any end-to-end BB security may be problematic and is left for future consideration.

3.3 Phase 2 - "Dynamic SLS Admission"

Phase 2 is referred to as "dynamic SLS admission" because, in addition to supporting inter-domain signaling as described in 3.2, BBs can dynamically set up new SLSs.

This includes the following:

- The granularity of SLSs is dynamic so that there can be one or many SLSs associated with each SLA. Many SLSs can be created from a single SLS
- SLSs can be destination (prefix) dependent. This allows local admission decisions to use an already established SLS to grant service over multiple domains without having to signal to downstream BBs as in Phase 1.
- BBs have support for configuring PHBs in their boundary routers (e.g., through SNMP, COPS etc) to enforce the SLSs. This is an intra-domain issue to be described separately. BBs have the capability to provide preemption of one data flow for another.
- BB may use the mechanisms of remarking packets in addition to simply dropping the packet. [Editor's note: I'm not sure this is valid given the Qbone architecture. This statement is include to capture the last discussion on the document.]

Thus, the same BB-to-BB request/response protocol as described in Phase 1 can be used. To obtain more general destination dependent SLs, a response may include a new (more general) destination prefix. The decision about establishing/granting a destination prefix dependent SLs is taken by the downstream BB. Prefix dependent SLs are useful for large and well provisioned destination subnets.

Additional support that can be provided is:

- BBs may have support to initiate reconfiguring internal routers so that the amount of aggregate resources available for a traffic class can be changed.
- BBs may have support for notifying an operator when there are insufficient resources in a domain or when police points are seeing excess traffic. This also requires notification from traffic conditioners in routers to the BB, which is an intra-domain issue that will be described separately.

3.4 Phase M- "Multicast"

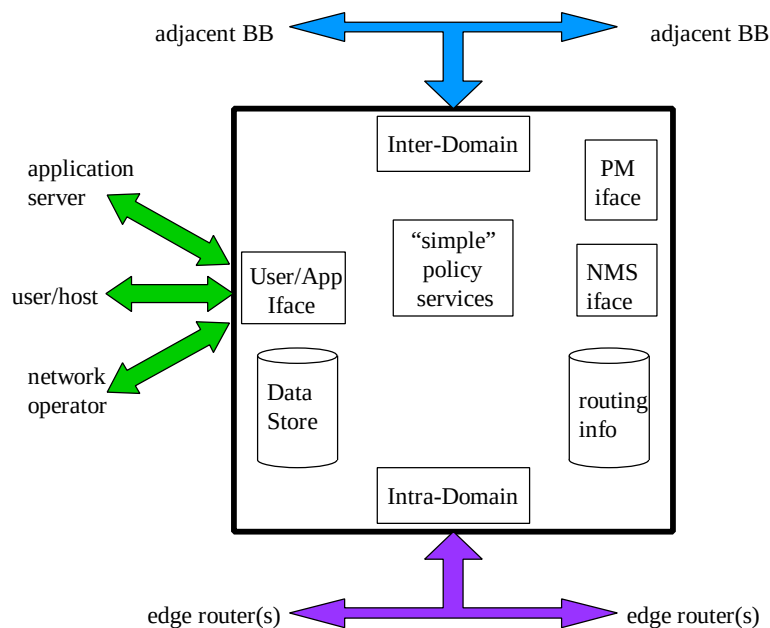
After the Bandwidth Broker works for unidirectional unicast, the Bandwidth Broker team will investigate the use of bandwidth broker for Multicast. Some of the interesting issues we will consider during the Multicast phase are:

- Shall we reserve bandwidth from the sender of a multicast group or from the receiver?
- What happens to the multicast aggregated flow if the end-to-end path does not support the full bandwidth.

We will add additional issues here as we consider multicast

4. BB Architecture

Having discussed some things a BB is and isn't and some functions it should perform, etc. We can now take a stab at putting together a general, functional level architecture for a BB. Not all components may be required by all implementations, but in general this architecture presents the necessary components to satisfy the specifications/requirements presented in the remainder of this document. (Again, notice the similarity to Policy Server architectures as seen elsewhere.)



A bandwidth broker can have the following general components, for which we discuss various specifications/requirements later:

User/application interface

- manual user/network manager GUI
- requests directly from user/app on end host
- requests from app server, e.g. H.323

Inter-domain communication interface

- between BBs in adjacent domains
- "negotiating" SLA information between domains

Intra-domain communication interface

- between intra-domain router and BB
- setting intra-domain device parameters for QoS/policy enforcement
- performing traffic management/ traffic engineering

Routing table interface

- BGP routing information for inter-domain
- Internal routing information for intra-domain
- Someday may utilize and influence QoS routing protocols

Data Repository

- used by all components
- may share with policy manager if PM exists
- may share with network management system (NMS) if NMS exists

Policy Manager interface (alternatively, BB may provide policy management services)

- utilize complex QoS/policy management functionality in policy manager
- coordination of service level contracts and network resources
- provide complete admission control processing
- AAA functions

Network Management interface (alternatively the BB may provide network management service)

- for coordination of network provisioning and monitoring
-

Editorial note: changes to this section include the following suggestions from Ruediger Geib

- 1) We need to clarify what is meant by "performing traffic management/traffic engineering" under the intra-domain communication interface. To Ruediger , this term implied a network management task. If traffic engineering is a network management task, then either state explicitly that BB and Network Management System is one unit or change the sentence to clarify that this is an interaction with the network management system.
- 2) Changes to Data Repository to include sharing data with NMS system
- 3) Changes to Policy Manager to include AAA functions
- 4) Changes to Network Management interface to suggest the BB may provide network management service

We will discuss this comment at the next BB on August 18th.]

5. BB Requirements

This section describes the requirements for a BB for deployment and interoperability in the Qbone. We focus our attention, in this version of the document, on those requirements necessary for Phase 0 and Phase 1 deployment (see Sections 3.1 and 3.2 above). Items which are not specific requirements for these phases may be mentioned but detailed descriptions of such items may be omitted until future versions.

5.1 *User/Application Interface*

- GUI for network operator to:
 - Populate data repository with simple network topology and configuration information
 - Input simple policy rules into the system
 - Manually input BB service requests to the system
 - Query the BB for existing service commitments, resource provisioning, etc.
 - Respond to operator with errors, etc.
- Host/User interface for:
 - Host/user applications to directly make BB service requests to the system
 - Users to query the BB for existing service commitments (granted to them); security required to prevent users from viewing commitments made to other users
 - Mechanism to respond to user/application
 - For user interface, same GUI (e.g. Java) may be used as for network operator; for applications running on hosts, protocol required
- Server/Gateway interface for:
 - Requests from application servers (e.g. H.323 gateway) on behalf of hosts/users
 - Probably not a "Phase 0" requirement
 - Can use same protocol as for user apps on hosts?

5.2 *Inter-domain Interface/Protocol*

A BB must respond to resource allocation requests received from either the Intra- or Inter-domain interface. During the course of processing these requests, the BB will communicate with DS edge devices and/or peering BB's. Edge devices will be configured as required to fulfill a request, and resource status and availability will be updated in affected domains.

This section defines the interface between adjacent cooperating Bandwidth Brokers (BB's), and specifies the mechanisms by which adjacent Bandwidth Brokers will communicate. Additionally, the specific data elements that will be communicated are identified.

It is pretty much assumed that the Inter-broker interface will take the form of a protocol. Note that it is not the intent of this document to define a protocol, or to recommend an existing protocol. That being said, you will notice that this section comes very close to defining a complete protocol. It turns out that this is a very effective method for capturing the interface requirements.

5.2.1 Operational Overview

The Inter-domain interface consists of different components. Based on the service definition the BB will administrate the resources available for each particular service.

The inter-domain interface will take the form of a request-ack based stateless protocol. BB's that wish to communicate must first establish a connection between pairs of adjacent BBs and then use these connections to exchange resource requests. Connections between BB's are always pair-wise, and bi-directional (either BB can request resources from the other). SLSs, however, are uni-directional; two SLSs need to be established - one in each direction. These comments on the protocol functionality do not imply any choice of a transport protocol to carry the BB information.

The general stages in the operation of this protocol are:

Local Configuration: The protocol loads local configuration information as it starts.
Session Initialization: Once running, BB's establish connections with adjacent BB's.
Service Setup: Connected BB's exchange requests for services for which they have SLA's in place.
Resource Allocation: BB's request resources from peers, and respond to resource requests.

Additional capabilities such as resource negotiation, administration and maintenance are provided and are also described below. Critical network infrastructure such as BB must tolerate faults within the infrastructure such as loss of individual BBs or connections between BBs. A section below also discusses these issues.

Some of these capabilities are mandatory, and others are optional. Unless otherwise specified, capabilities and requirements denoted as mandatory must be included in any implementation. Optional capabilities may be required to support advanced inter-domain communication that is likely to occur during the later stages of BB deployment (see section 3 – Deployment).

The following sections describe the specific requirements for, and operation of, each of these stages. Note that X, Y and Z will be used to denominate adjacent DS domains which are negotiating differential IP services. The specific data elements included in each type of message are defined in section 4.2.2.

5.2.1.1 Local Configuration

It is expected that the definition of a service between two domains and the initial resources available to support this service be defined without BB communication. Besides the definition of the resources, other service specific parameters like DS code points, PHBs and treatment of out-of-profile traffic must be agreed to before inter-domain communication can begin.

In the following, it is anticipated that a bilaterally agreed DS code point identifies a service between two DS domains. Networks should use the DSCP's suggested by IETF for services based on any per hop behaviors standardized by the IETF.

All BB's must have access to the following information concerning the Inter- and Intra-domain connections of each DS boundary router in their domain:

- a) Service-specific CAC (including relations of resource consumption of administered service to resources available for other services). [editorial note: Gary Hoo inquires: Is this a BB function only? How does this impact the router. Editor's example might be: Two domains exchange traffic over an OC-12 link. The OC-12 link has 60% of the total capacity for traffic allocated to QoS traffic. Of that 60% of the link reserved for QoS, 50% of the QoS traffic has already been allocated to SLA #2 who can go to a grand total of 70% of the total QoS bandwidth. CAC include the BB ability to know these factors prior to allocating the traffic. (Is this a valid example.) We will discuss this at the 8/18 meeting.]
- b) Maximum available service-specific resources.
- c) Service-PHB mapping (including Intra-domain to Inter-domain PHB mapping).
- d) Service-DSCP mapping (including Intra-domain to Inter-domain DSCP mapping).
- e) SLA ID.
- f) Methods of signaling allowed by the SLA. [editor's note: Gary Hoo inquires: "what does this mean to a router? The Editor's understanding is the methods of signaling can be end-to-end or immediate as described in sections 3.2.3. Optional parameters for this signaling are the mode of routing BB packets (section 3.2.4)]

All items (a through f) are mandatory.

5.2.1.2 Session Initialization

Before any services can be configured between two adjacent domains, the BB's will have to connect and establish communication. During this initialization process, a BB communicates:

- a) AS(es) of it's own DS domain
- b) BB address/ID of sending BB
- c) List of adjacent BB's

All of these items are mandatory.

The list of adjacent BB's may support:

- 1) additional checks for multiple BBs interconnect between domains
- 2) BBs immediately adjacent to BB in the remote domain. The adjacency can either be physical or logical.

In an environment where multiple BBs are deployed at the Inter-Domain exchange points, two domains may have multiple connections between each domain. A highly inter-connected mesh such as figure 3 demonstrates provides redundancy in the face of failure. The list of BBs sent by BB1 to Domain 2 indicates the BBs that BB1 is attached to. The list would include: BB2, BB4 and BB5.

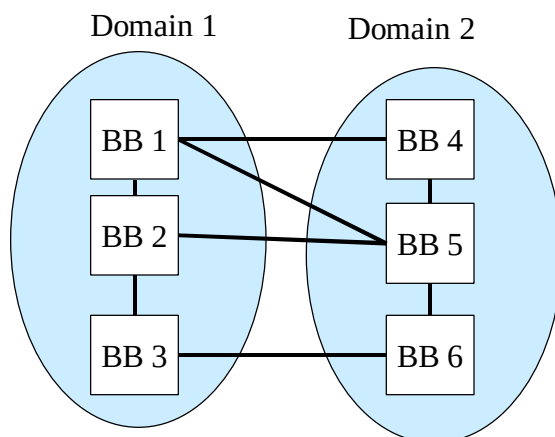


Figure 3: Highly Meshed Interconnection between two domains

5.2.1.3 Service Management

BB's are responsible for managing the services that are available to customers via their Inter and Intra-domain interfaces. Services are agreed to in SLA's, and the defined by the technical content of the associated SLS's.

The capabilities required for managing SLS defined services include:

- Service Set-up (mandatory)
- Service Cancellation (mandatory)
- Service Re-negotiation (optional)

A. Service Set-up (mandatory)

Prior to a service being set up between two domains, the peering BB's must be configured in a consistent manner with regard to the agreed to bilateral service. All resources required to support the service should be provisioned and "reserved" in each local domain. When the necessary resources are in place and defined to each BB (see section 5.2.1.1 – Local Configuration), Service Set-up Request messages can be generated.

The SSR message is simply a mechanism for a BB to confirm it's SLA(s) with its peering BB, and ensure that the peer has the resources in place to honor the SLAs.

Editor's note: The text from here on is inconsistent with the text in sections 2-4. In order to make the text consistent, if the SSR deals with the SLAs the RAR function below has to be broken into an SLS function and a RAR function. We will discuss this break-out in the 8/18 BB meeting. The proposed text for the SLS function is included below and the edits in the next sections reflect that set of concepts.]

A "Service Set-up Request (SSR)" message can be initiated in a number of ways:

- On command from the network operator
- Automatically at system start up or restart (based on allocated services which had previously requested and which have not expired or been cancelled)

The SSR message contains the following data:

- a) BB_ID of the BB sending the SSR
- b) SSU_ID (Service Set Up – ID)
- c) SLA_ID
- d) PHB - value used should be 16 bit PHB identifier specified in QoS Draft
- e) DS_egress_router [editorial Q: Is this the DS Egress router that traffic will flow as it exits the DS domain the BB is contained in? My understanding is the DS path has only 1 Egress router per resource allocation. Is this true when we reach high aggregation points? Can we discuss this point at the next BB meeting (8/18)?]
- f) Adjacent_DS_downstream_router [editorial Q: Is the adjacent DS downstream router the router within this BB's domain that is the ingress to the DS domain that the remote BB is in? Or is this the remote Domain's router. Again, we will need to discuss this at the next BB meeting (8/18)].
- g) DSCP (multiple possible)
- h) DSCP_specific_resources_downstream (one per DSCP)
- i) Flag on the type of signaling - End-to-end or immediate

Note that this request is for unidirectional service by default (see [QBONEARCH]).

[Editor's note: The SSR request assumes 1 SLA between two BBs per SSR. Can two diff-serv domains have two SLAs between the same two domains? If so, we need to discuss the addition of concept and text here to deal with the multiple SLAs between domains.]

The receiving (downstream) BB will check consistency with the SLS information stored in it's own database. If it is consistent, the BB will reply with a "Set-up Answer (SA)".

The SA message contains:

- a) BB_ID of responding BB
- b) SSU_ID
- c) SLA_ID
- d) Confirmation/Reject Indication
- e) Flag on type of signaling remote side uses

Each BB must keep track of all resources that are active, or allocated to a specific service. The status of resources must be updated when a SA confirmation is received.

Note that the service may not be symmetric, i.e. the DSCP_specific_resources X to Y may differ from those Y to X.

B. Service Cancellation (mandatory)

If a SLA is terminated, the service must be cancelled.

The "Service Cancellation (SC)" message contains:

- a) BB_ID
- b) SSU_ID
- c) SLA_ID

This request may be initiated by a network operator command or triggered by an expiration date. A SC will be checked for consistency on reception and replied with a "Cancellation Answer (CA)" message.

The CA contains:

- a) BB_ID
- b) SSU_ID
- c) SLA_ID

When the Cancellation Answer is sent the BB updates the status of all resources that were reserved for the specific service. When the adjacent domain receives a Cancellation Answer, the relevant conditioning parameters of the adjacent DS boundary routers are disabled.

Note that due to the unidirectional nature of the service set-up a service may be canceled in direction X to Y while it stays operative in direction Y to X.

C. Service Re-negotiation (optional)

A SLA may allow a BB to re-negotiate an existing service. This is achieved with a "Service Renegotiate (SR)" message. SR messages are initiated:

- Manually via a network operator command
- Automatically by the BB based on some service specific threshold which is exceeded

If additional resources are to be added for a service, the status of the newly allocated resources must be updated in the BB's database.

The SR is confirmed by a Set-up Answer message. On confirmation the reserved resources are deemed active in both domains, and the service is available for use. Both BB's must update the traffic conditioning or policing parameters of the affected DS boundary routers in their domains.

The SR contains:

- a) BB_ID
- b) SSU_ID (the same as the earlier Service Set-up)
- c) SLA_ID
- d) DSCP_specific_resources_downstream (new value)

All other parameters are expected to be the same and are not re-transmitted.

5.2.1.4 Resource Allocation

[Editorial note: This section has have been added to attempt to make the message formats consistent with sections 2-3. We will discuss these additions in the 8/18 meeting. These additions are included for discussion and may be removed after the 8/18 discussion.]

After completing the initial service set-up procedure, the service is available. Note that the service has not actually been allocated to any specific flows. The BB should not allow this available service to be used until it receives Resource Allocation Requests (RARs) for the service within a Domain and SLS between domains.

The RAR requests within a Domain allow a user or network application to request resources from the Bandwidth broker. The RAR allow BB's in a single domain to keep track of the actual quantity of an available service that is actually allotted to traffic within a domain, and to make admission decisions based on this. The BB is able to process RARs from within a domain once the SSR has been confirmed.

The SLS requests allow BB's in adjacent domains to keep track of the actual quantity of an available service that is actually allocated to traffic, and to make admission decisions based on this. The BB is able to process SLS once the SSR has been confirmed an inter-domain BB the BB is communication with..

Note that the procedures described below define a unidirectional resource allocation. Bi-directional resource allocations require that appropriate reservations be requested in both directions for the relevant connection.

Resource allocation involves request and response messages, which have been named Resource Allocation Request (RAR), and Resource Allocation Answer (RAA), respectively.

A. Resource Allocation Request (mandatory)

A BB receives a "Resource Allocation Request (RAR)" message either from the Intra-domain BB. In either case a similar set of checks will be performed. As long as the requested resources are available, the Intra-domain RAR processed by BB X results in a RAR to the adjacent BB Y. If BB X received a RAR from BB Y, it will return a "Resource Allocation Answer (RAA)" confirming the requested availability of resources through domain Y.

The Resource Allocation Request (RAR) consists of:

- a) BB_ID
- b) SSU_ID (valid SSU_ID of the service set up)
- c) RAR_ID
- d) DS_egress_router_ID *[editor: Is this necessary? - Olov]*
- e) destination prefix
- f) Allocation Period (Start-Time and End-Time)
- g) DSCP (multiple possible)
- h) DSCP_specific_resources_downstream (one per DSCP)
- i) Optional parameters

Based on bilateral agreements, the following optional parameters may be part of the RAR:

- o) Source address, destination address, protocol, port numbers, profile information, negotiable parameters, time-dependant different reservation schemes.

Note: The mandatory parameters are used to check the availability of resources and to update conditioning parameters of aggregates. The optional parameters may support fine grain or per-flow conditioning for all or parts of the traffic between two DS domains.

B. Resource Allocation Answer (RAA)

The Resource Allocation Answer consists of:

- a) BB_ID
- b) SSU_ID
- c) RAR_ID
- d) Confirmation/Reject/Progress Indication

The following processing occurs when RAR messages are received (this describes processing of mandatory parameters only).

C. Service Level Specification Request (inter-domain) SLSR

A BB receives a “Service Level Specification Request (SLSR)” message either from the Inter-Domain BB. In either case a similar set of checks will be performed. As long as the requested resources are available, the Intra-domain RAR processed by BB X results in a RAR to the adjacent BB Y. If BB X received a RAR from BB Y, it will return a “Service Level Specification Acknowledgement (SLSAck)” confirming the requested availability of resources through domain Y.

The Service level Specification (RAR) consists of:

- a) BB_ID
- b) SSU_ID (valid SSU_ID of the service set up)
- c) SLS_ID
- d) DS_egress_router_ID *[editor: Is this necessary? - Olov]*
- e) destination prefix
- f) Allocation Period (Start-Time and End-Time)
- g) DSCP (multiple possible)
- h) DSCP_specific_resources_downstream (one per DSCP)
- i) Optional parameters

Based on bilateral agreements, the following optional parameters may be part of the SLSAck :

- o) Source address, destination address, protocol, port numbers, profile information, negotiable parameters, time-dependant different reservation schemes.

Note: The mandatory parameters are used to check the availability of resources and to update conditioning parameters of aggregates. The optional parameters may support fine grain or per-flow conditioning for all or parts of the traffic between two DS domains.

D. SLSAck

The Resource Allocation Answer consists of:

- j) BB_ID
- k) SSU_ID
- l) SLS_ID
- m) Confirmation/Reject/Progress Indication

F.) Processing of SLS and RAR messages

The following processing occurs when an RAR or SLS messages are received (this describes processing of mandatory parameters only):

RAR received (by BB Y within a domain) from the User-interface or an upstream BB X with destination domain Z

The BB at Y processes the RAR as follows:

1. If applicable: map Intra-domain RAR parameters to Inter-domain SLS parameters according to the SLA with domain Z. Send a RAR to other Intra-Domain BBs.
2. Check availability of resources at the inter-domain link by running the service specific CAC.
3. If resources are available, reserve as requested.
4. If within a domain, Send a RAR to the relevant downstream BB Z. If downstream is outside this domain, send an SLS to the downstream domain.

5. If the downstream BB receiving either the SLS or RAA from BB Z confirms availability of resources, put the reserved resources on active state.
6. If applicable: configure the shaping parameters of the DS egress router to domain Z with the new value of the allocated resources.
7. Send a RAA to the upstream BB X interface or to the User-interface.

SLS received (by BB Y) from BB X on the Inter-domain Interface

The BB at Y processes the RAR as follows:

1. If applicable: map Inter-domain SLS parameters to Intra-domain RAR parameters to be sent to other BBs within the domain.
2. Check availability Intra-domain resources by running the service specific CAC.
3. If resources are available, reserve as requested.
4. if the destination is not the own domain (called Z), process the SLS and generate the correct SLS to send to the next domain. Send the appropriate RAR to other BBs within this domain. If the destination is within this domain, progress SLS for this user-domain otherwise.
5. If the) SLSack received from the BB in the next domain and the RAA within this domain confirms availability of resources, put the reserved resources on active state.
6. Update the policing of the DS egress router with the new value(s) of the allocated resources.
7. Send a SLSack to the upstream domain BB Y.

If sufficient resources are not available a RAA with a reject indication will be sent to BBs within this domain and a SLSack will be sent to the BB in the upstream domain. In this case, the reservation of resources will be released.

Depending on the mode of operation, a domain may either directly answer a SLS with a SLSack indicating confirmation or it may indicate progress only. In the latter case, the BB receiving a "progress" information knows that his SLSack was received and the relevant request is forwarded to the destination. In this case, the BB receiving the "progress" indication may start a timer to allow further domains to process the SLS.

5.2.1.5 Session Maintenance

In addition to resource and service request messages, any protocol which implements the Inter-domain interface must also perform a variety of session maintenance activities. These include:

1. The protocol could make use of an existing connection oriented transport protocol, such as TCP (optional)
2. Keep Alive procedure (mandatory)
3. A mechanism to cope with BB outages or loss of communication between BB's is required. Issues to be handled include:
 - A BB which has re-started must reconstruct its own state (mandatory)
 - A recently re-started BB must inform its neighboring BB's of its state (mandatory)
 - A recently re-started BB is responsible for synchronization/re-negotiation of services and resource allocations with adjacent BB's (mandatory)
4. The interface must include timers to trigger repeat requests in the case of missing or delayed answers (mandatory)
5. If a set up request is rejected or no "Answer" message is received after repeated request attempts, an alarm should be sent to the network operator of the requesting BB (optional)

6. There must be a defined behavior governing the processing of unexpected messages (e.g. answers without requests) and unexpected parameters in messages (mandatory)

5.2.2 Minimum Performance Capabilities

To be determined.

5.2.3 Future Enhancements

1. An optional change of the procedures to support bi-directional resource allocation.
2. Specification of desired operational characteristics when end to end signaling is to be deployed.

5.3 *Intra-domain Interface/Protocol*

This section defines the interface between a Bandwidth Broker and the routing equipment in the local domain (typically edge routers). It should be stated here that the *method* of communication between routers and BB's in a local domain is only of local concern within a domain. Within a domain the intra-domain protocol must provide a common means by which to allocate resource within the domain. These resource allocations will also need to be communicated to the routing equipment within the domain.

This section does not specify any specific method of communication, as this may vary based on the equipment interfaces supplied by the various vendors.

What this section will specify is the types of requests that BB's must be able to make to routers, and the data that will be most likely contained in these requests. Note that the exact data may vary based on vendor specific DiffServ implementations within routing equipment.

5.3.1 Interface Mechanisms

It is pretty much assumed that the Intra-domain interface will be implemented as some automated process. It is through this automated process that the BB will configure (or reconfigure) routers to enable or allocate various DiffServ services. Any automated process will have to rely on the router configuration mechanisms supplied by the router manufacturers. The following list describes mechanisms, which exist today (although not all routers support all mechanisms).

1. COPS

Routers, which include a COPS-Provisioning client [COPS-PROV], can be configured via the COPS protocol. The BB would include a COPS server component and would use the protocol to request that the various DiffServ services be turned on and/or off.

2. SNMP

Routers, which support SNMP SET messages, and publish a MIB for their DiffServ components, could be configured via SNMP.

3. Telnet

Most routers supply command line interfaces (CLI's). A BB could communicate and configure a router by automating a telnet session, and executing the appropriate configuration commands via the CLI.

5.3.2 Intra-domain Interface Capabilities

The intra-domain interface should allow the BB to manage the DS services and service allocations within any router which routes DS packets. This must include edge routers, and may include other (non-edge, or internal) routers.

The management of a DS router involves the configuration and control of the various Traffic Conditioning (TC) components within the router. Specific details concerning TC components are included in [DIFFEDGE] and [RFC2475], and are listed here for completeness. TC components include:

- Classifiers
- Meters
- Markers
- Shapers
- Droppers

The data elements required to configure these components is listed in the following paragraphs.

1. Classifiers

DiffServ supports DS and MF classification.

DS classifiers require: DSCP.

MF classifiers require one or more header fields, such as source address, destination address, DS field, protocol ID, source port and destination port numbers, and other information such as incoming interface [RFC2475].

2. Meters

3. Markers

4. Shapers

5. Droppers

5.4 Routing (Table) Interface

When processing BW requests (or whatever they're called), BGP routing information may be accessed, along with policy management information, for making admission control decisions:

- AS numbers
- Preferred path(s) into neighboring domains
- Resource availability information on egress links

5.5 Multicast

[TBD - Sue Hares to provide text.]

6. References

[BCITBB] R. Neilson, *CA*net II Differentiated Services Bandwidth Broker Specification*, October, 1998.

[TWO-TIER] F. Reichmeyer, L. Ong, A. Terzis, L. Zhang, and R. Yavatkar, *A Two-Tier Resource Management Model for Differentiated Services Networks*, IETF draft-rotzy-two-tier-01.txt, November, 1998.

[AGENTS] O. Schelen and S. Pink, *Reservation Agents Paper?*

[RFC2475] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, *An Architecture for Differentiated Services*, IETF RFC 2475, December, 1998.

[RAPFRAME] R. Yavatkar, D. Pendarakis, and R. Guerin, *A Framework for Policy-based Admission Control*, Internet Draft, < draft-ietf-rap-framework-01.txt>, May 1998.

[QBONEARCH] Internet2 QoS Working Group Draft, *Draft Qbone Architecture*, May, 1999.

- [COPS-PROV] F. Reichmeyer, K. Chan, D. Durham, R. Yavatkar, S. Gai, K. McCloghri, S. Herzog, A. Smith, *COPS Usage for Policy Provisioning*, IETF drfat-sgai-cops-provisioning-00.txt, February, 1999.
- [DIFFEDGE] Y. Bernet, D. Durham, F. Reichmeyer, *Requirements for Diff-serv Boundary Routers*, IETF draft-bernet-diffedge-01.txt, November, 1998.