Internet Engineering Task Force Working Group
Internet Draft
draft-pan-bgrp-framework-00.txt
January 14, 2000
Expires: July, 2000

P. Pan, E. Hahne, H. Schulzrinne
Bell Labs/Columbia U.

BGRP: A Framework for Scalable Resource Reservation

STATUS OF THIS MEMO

Abstract

Resource reservation needs to accommodate the rapidly growing size and increasing service diversity of the Internet. This memo first defines the scaling problem in today's Internet backbone, and briefly discusses several existing resource management approaches. Then we will present a distributed approach and introduce a protocol, called the Border Gateway Reservation Protocol (BGRP), for inter-domain resource reservation that can scale in terms of message processing load, state storage and control message bandwidth.

The main idea of our approach is to build a sink tree for each domain network. Each sink tree aggregates reservations from all data sources in the network. Sink tree initiation, maintenance and termination involve only backbone border routers. Within each domain, the network service providers manage network resource and direct user traffic independently. At the border routers, the service providers can use BGRP to setup domain-level reservation trunks base on bi-lateral agreement. Since routers only maintain the sink tree information, the total number of reservation states at each router

scales, in the worst case, linearly with the number of domains in the Internet.

For bandwidth reservation, BGRP relies on differentiated services for data forwarding. As a result, the number of packet classifier entries is small. To reduce the protocol message traffic, routers may reserve domain bandwidth beyond the current load so that sources can join or leave the tree or change their reservation without having to send messages all the way to the root for every such change.

1 Introduction

Resource reservation had originally been defined to support end-to-end QoS guarantees for a range of QoS-sensitive applications including multimedia-on-demand and teleconferencing. Recently, service providers have started to use the same reservation mechanisms to provide customer-level VPNs and to dynamically provision network resource. To support this, the RSVP [1] has been modified [2] to carry MPLS information [3,4] to setup LSP's (Label Switched Path) across the Internet. Similarly, it can also be used to set up optical crossconnects (OXCs) for optical devices [5].

Hence, resource reservation schemes must scale well with the rapidly growing size of the Internet. A router may be able to handle tens of thousands of simultaneous reservations [6], but not hundreds of thousands, and certainly not millions. Today's traffic volume is bad enough: as we will show in Table 1 of Sec. 2 below, we have measured hundreds of thousands to millions of flows at the MAE-West network access point; if many of these flows were to request resource reservations, the protocol overhead would swamp the router. But projected future traffic growth is an even more serious problem. The overhead of the current protocol RSVP [7,1] grows like N**2, where N is the number of Internet end hosts. Data in [8] shows the growth of N over the last six years, from 2 million to 60 million. This means that N**2 grew from 4x(10**12) to 4x(10**15) during that time! With no end in sight, N**2 is growing much faster than improvements in processing speeds or memory sizes.

Therefore, we will have to find a reservation scheme that scales better than conventional RSVP. In this paper we will propose a protocol, called the Border Gateway Reservation Protocol (BGRP), that fixes this scaling problem in two ways. First, BGRP overhead scales linearly with the size of the Internet; i.e., N**2 is reduced to N. Second, BGRP uses "a smaller N". The overhead of the basic BGRP protocol is proportional to the number of Internet carrier domains (also called Autonomous Systems (AS)), while an enhanced version of

BGRP has overhead proportional to the number of IP networks (i.e., the number of announced IP address prefixes).

2 Problem Definition

What exactly is the resource reservation "scaling" problem? Since resource reservation is not yet widely used, we have to extrapolate the likely volume of reservation state from other observations. To that end, we collected a 90-second traffic trace from the MAE-West network access point (NAP). We categorized about 3 million IP packet headers [9] from June 1, 1999, according to their transport-layer port, IP address, IP network prefix and BGP Autonomous System (AS) number. Table 1 shows the results; for example, if we use RSVP, the total number of reservations can range from 20,857 if we reserve source-destination AS pairs up to 339,245 if every flow identified by a unique 5-tuple gets its own reservation.

Granularity	flow discriminators	flows
Application	source address, port	143,243
	dest. address, port, proto.	208,559
	5-tuple	339,245
IP Host	source address	56,935
	dest. address	40,538
	source-dest. pairs	131,009
Network	source network	13,917
	dest. network	20,887
	source-dest. pairs	79,786
AS	source AS	2,244
	dest. AS	2,891
	source-dest. pairs	20,857

Table 1: Flows and aggregations based on a 90-sec packet trace from MAE-West

Can network routers handle hundreds of thousands of reservations? After all, telephone switches handle tens of thousands of simultaneous calls. In a recent study [6], we showed that a low-end router can set up 900 new RSVP reservations or maintain up to 1600 reservations per second, allowing it to sustain about 45,000 flows. To sustain that rate, the router has to suspend routing computation and packet forwarding due to its hardware and CPU constraints. While

backbone routers have more CPU power than the low-end router used for the measurements, other results [10] indicate that frequent routing computation due to route instability may already tax the CPU. Thus, we, along with some RSVP developers we talked to, believe that in many networking environments, routers do not have enough CPU power to sustain hundreds of thousands of reservations. Developers from several high-end router vendors have acknowledged that RSVP processing could consume between 30% to 50% of router CPU cycles.

From Table 1, we observe that there are about 21,000 unique source-destination AS pairs which a backbone router should be able to handle. However, this number is artificially low due to the small 90-second window. Over the span of a month (May 1999), MAE-West saw 4,908 unique source ASes, 5,001 unique destination ASes and 7,900,362 unique AS pairs, out of the 25 million possible combinations. The measurement result is consistent with Bates's statistics [11]. Thus, unless edge routers tear down AS-level trunks frequently, there may be too many AS-level reservations to sustain in backbone routers.

The table also indicates that the number of source and destination AS's and networks is relatively small. Bates' statistics[11] show that there were approximately 5,000 autonomous systems and fewer than 60,000 network prefixes in the Internet in June 1999. Hence, if we set up reservations based on either source or destination AS's or network prefixes, we can readily keep the reservation count at levels sustainable by today's routers.

Today's inter-domain routing protocol, BGP [12], establishes "virtual edges" by using reachability as a definition for the existence of a link in the graph. In case of paths of equal weight, current practice dictates that the router forward all packets over only one path. This practice guarantees that routing always follow sink trees. Hence, if reservations are make along routes chosen by the BGP routing algorithms, it is natural to aggregate these reservations along the sink trees formed by routing. We will discuss a sink-tree reservation approach in detail in Section 6.

2.1 What are we reserving?

It has been argued that since network link bandwidth is finite, it is unlikely that a link would see thousands of reservations. On the one hand, if we assume that a voice call is the finest granularity of reservation, an OC-192 link carrying 16 kb/s voice flows could support up to 600,000 such calls. On the other hand, our traces indicate that the MAE-West link carries only several hundred high-volume flows[13]. The latter seems to lead to a conclusion that reserving resources for high-volume flows does not pose scaling problems.

However, with the deployment of IP telephony, the number of real-time traffic flows is likely to increase. At the same time, network bandwidth will likely rapidly increase due to the deployment of optical switches into the Internet backbone. Both factors may result to a large number of bandwidth reservations in the network.

More importantly, some of the recent IETF activities on MPLS and Traffic Engineering[14] have suggested of using the reservation signaling protocols to set up MPLS LSP (Label Switched Path) tunnels and to provide service differentiation among users. MPLS LSP's have very similar characteristics as ATM VP's. Each LSP connects two network nodes and the connection in turn carries datagram traffic. We envision that it is quite possible to have many LSP's on a given link in the backbone.

In conclusion, we believe that reservation signaling protocols must be able to carry control information for QoS (such as bandwidth) reservation and for MPLS label setup.

3 Related Solutions

Recently, several authors have addressed scalable resource reservation, using either a server-based or a router-based approach.

In server-based approaches, each domain has a bandwidth broker (or agent) which is responsible for selecting and setting up the aggregated reservation sessions. This approach has the advantage of removing the message processing and storage burden from routers. However, synchronizing reservation information among the bandwidth brokers and the routers may be complex. No aggregation takes place, so that each broker still has to deal with the requests of individual flows. Also, care has to be taken so that the broker does not become a single point of failure for the domain. Variations of the server-based approach have been described by Blake et al. [15], Schelen and Pink [16], Berson et al. [17], and Reichmeyer et al. [18]. The latter proposal suggests a two-tier system where, within each domain, hosts use intra-domain reservation protocols such as RSVP to set up reserved flows. Inter-domain reservation protocols set up coarsely-measured reserved flows between domains. However, the proposal leaves the actual mechanism undefined.

Awduche et al. [2], Guerin et al. [19] and Baker et al. [20] have proposed a router-based approach by modifying RSVP to support scalable reservation. (Awduche's LSP tunnels [2] are designed to support intra -domain traffic engineering, but may also be used to set up trunks crossing multiple domains.) These proposals aggregate per-application reservation requests into reservation "trunks" between pairs of domains, by modifying sender template and session

objects in RSVP to carry address and mask ("CIDR blocks") or autonomous system (AS) numbers instead of 5-tuples (sender IP address, sender port, receiver IP address, receiver port, protocol). However, this implies that routers in the backbone may have to maintain reservation state proportional to the square of the number of CIDR blocks or autonomous systems. Since the number of AS is currently about 6,000, the number of AS pairs exceeds 36,000,000. As we argued in Sections 2, this is excessive. A more aggregated reservation scheme is needed.

Finally, Feher et al. [21] have recently proposed a stateless reservation mechanism called Boomerang, where end users send reservation requests and refresh messages to set up and maintain reservations. No per-flow state is stored at routers. However, the scalability of the control message processing is an issue.

4 Terminology

We define the following terms in the memo.

Domain: The term "domain" has the same meaning as the one being used in BGP [12]: each domain has an unique AS (autonomous system) number, and can exchange user traffic with its peers. Each domain is under a single common administration.

Border Router (BR): A domain connects to a number of other domains via border routers. We assume all border routers use BGP4 for inter-domain routing. For simplicity reason, we only consider the EBGP (External-BGP) border routers at present time.

Downstream and upstream: We define the directional terms "upstream" and "downstream" with respect to the direction of data flow. The traffic direction from source to destination is "downstream"; destination to source, "upstream".

Reservation sender and receiver: A reservation sender is an upstream border router that originates reservation messages. A reservation receiver is a downstream border router that terminates reservations.

Reservation aggregation: Reservation aggregation occurs if multiple reservations coming from different sources but going toward the same destination can be "added" together to create a single reservation.

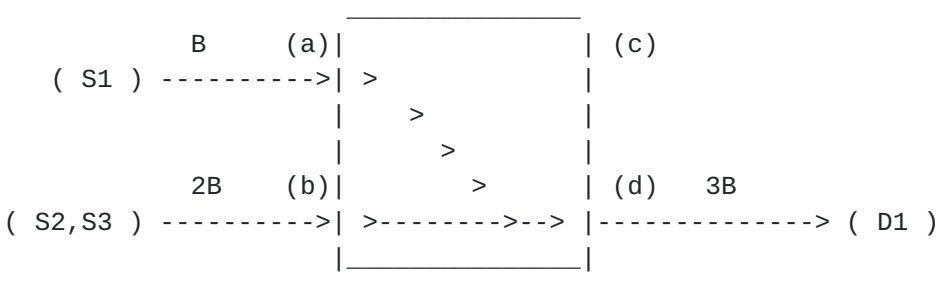

Figure 1: Bandwidth reservation aggregation.

Figure 1 illustrates the concept of bandwidth reservation aggregation. (a), (b), (c) and (d) are the interfaces for a router. Reservations from data source S1, S2 and S3 are all going to destination D1. In the example, each source needs (B) units of bandwidth. At (d), all individual reservations are aggregated together. The amount of reservation at (d) is (3*B) as a result.

5 Requirements and Assumptions

Message processing: The cost of processing reservation messages depends on the complexity of handling each message and the frequency of reservation messages. Instead of setting up reservations across domains as application flows arrive, we rely on pre-computed reservations made in advance. This also implies that each reservation would last for relatively long period of time (typically hours or days).

State storage overhead: Routers need to store reservation control information and packet classifier tables. To reduce the former, we propose a scheme for aggregating reservation flows. To reduce the latter, we rely on diff-serv [15] to eliminate per-flow queuing and processing, so that the number of queues is likely to be no larger than a few dozen.

Bandwidth overhead: The bandwidth consumed by setting up reservations should be small compared to the link bandwidth, both in steady state and with routing transients. This bandwidth overhead is typically proportional to the number of flows kept in routers, and thus minimizing the state storage overhead also helps here.

Intra-domain vs. inter-domain reservation: It is desirable that each domain can manage its own network resources and enforce its own internal traffic engineering policies. This implies that a domain only reveals simple delivery commitments to its peering domains in terms of bi-lateral agreement. The inter-domain reservation then uses these delivery commitments to establish a reservation path through multiple domains. Each domain sets up transit reservation flows using its preferred intra-domain reservation mechanism.

Routing interface: Reservation protocols do not do routing and only rely on the routing information to set up reservations along the data forwarding path. More importantly, reservation protocols must not interfere with route aggregation and effect routing protocol scaling properties.

6 Architecture

The Internet backbone consists of a number of domains, each of which has at least one border router (BR). Through BGP4, each BR learns about the other BR's within its own domain, and the directly connected BR's in the adjacent domains. Through out-of-band means, the BR's know of bi-lateral (or multi-lateral) agreements with the peering domains. Figure-2 illustrates a network structure.

Typically, the bi-lateral agreement specifies the inter-carrier policy information such as route filtering and route preference [22]. In the future, we envision that the bi-lateral agreement may also include policies for QoS guarantees between peering domains. It's worth noting here that the bi-lateral agreement applies only between two adjacent domains, and users do not always have the knowledge or the guarantees on which downstream domains their traffic will traverse through.

As being advocated in the IETF Traffic Engineering Working Group, the ISP's set up border-to-border (or edge-to-edge) intra-domain "virtual" trunks between border routers. At each NAP (Network Access Point) or POP (Point-of-Presence), the ISP's set up similar "virtual" trunks to interconnect with external domains. The goal here is to optimize the use network resource and traffic performance. Example of "virtual" trunks, as represented by the lines between BR's in Figure-2, includes Frame Relay, ATM, MPLS and DiffServ.

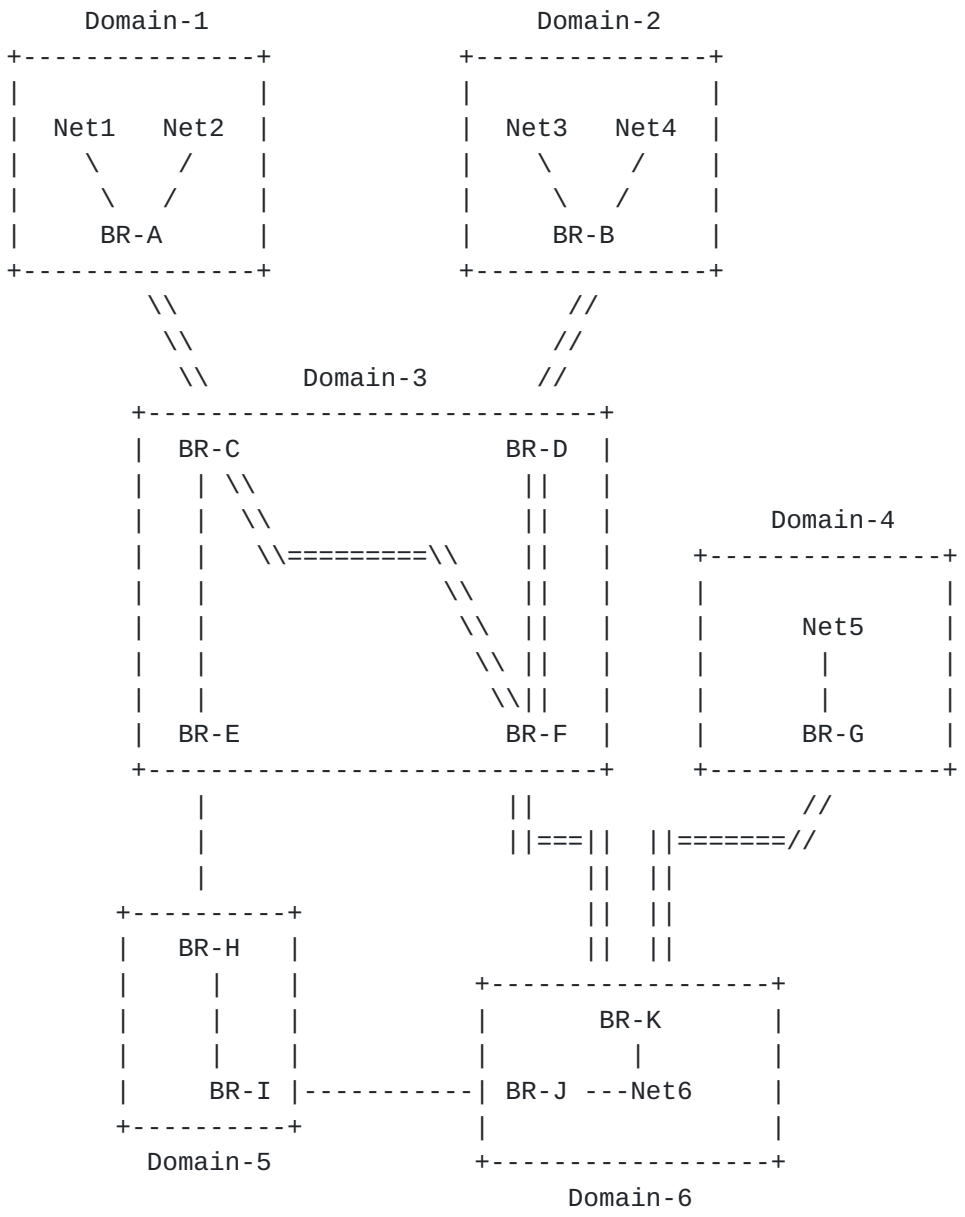

Figure-2: An inter-domain network example.

6.1 Sink Tree Model

Current BGP practice guarantees that routing paths form sink trees. In the example, the border router BR-K advertises the Net6 reachability to the rest of the network. Let's assume that per routing policy, BR-C always favors the routes advertised by BR-F. Thus all data traffic going to Net6 forms a sink tree with branches

(BR-A, BR-C, BR-F, BR-K), (BR-B, BR-D, BR-F, BR-K), and (BR-G, BR-K). The sink is BR-K.

To create inter-domain reservations, one obvious choice is to build reservations along the sink trees. This approach has a very desirable scaling property: the routers only need to remember the reservation "sinks". We can further improve the scalability by allowing only the BGP border routers to participate in the reservation process.

6.2 How to Create Sink Trees?

There are many ways to establish reservation sink trees inside the backbone.

One solution is the following: reservation senders can always query a central database somewhere to get the precise routing path prior to each reservation. A similar idea has been introduced in the RSVP LSP extension[2], in which an Explict-Route Object (ERO) is computed or queried by the reservation senders before each request. In addition, we had proposed a router-server tunnel data exchange mechanism using COPS [23]. This is a reasonable solution for intra-domain resource management, but it may not be applicable in inter-domain environment since the centralized database can be very difficult and costly to maintain and can have security implications.

Another solution is to allow the routing protocols, in this case, BGP4, to setup sink-trees at route advertising and RIB (Route-Information-Base) processing time. However this approach may effect route aggregation and cause routing scalability problems.

Here we propose a two-phase distributed reservation solution. The reservation senders send querying messages to the network. The queries will follow the data path and are delivered between BGP hops. At each BGP-hop, a routing path is selected based on the bi-lateral agreement. However, the routers do not pin down the reservation path and do not store the query data.

The reservation receivers keep track of all the queries, and construct sink-tree graphs from the information. The receivers send reservation request messages upstream to set up the actual reservations. At each border router, the router aggregates the reservations for each sink tree.

In the next section, we will describe a protocol, BGRP, based on the above ideas.

7 BGRP: Border Gateway Reservation Protocol

The Border Gateway Reservation Protocol (BGRP) is an inter-Autonomous System reservation protocol. It is to set up aggregated reservations over multiple Autonomous Systems (AS). The reservation originators and the terminators are at the BGP-speaking border routers. Reservation aggregation is at the AS level. Only BGP-speaking routers in the network participate in reservation process.

All traffic destining to a particular AS can be aggregated in a sink-tree fashion. Since backbone routers only maintain the sink tree information, the total number of reservations at each router scales, in the worst case, linearly with the number of AS's in the Internet.

BGRP runs over TCP and thus eliminates the need to implement fragmentation, retransmission and sequencing. BGRP uses "soft-state" to manage reservations, i.e., periodic refresh, to protect against events such as link failure.

BGRP is not involved in setting up and managing intra-domain reservations. We envision that the ISP's may use RSVP Traffic Engineering extension, LDP [24] or other means to manage internal network resource. BGRP only becomes useful when reservations need to be established across multiple AS's.

The current version of BGRP does not provide support for multicast traffic. This is because, first, the majority of the Internet traffic is unicast. It is not likely to change in the near future. Second, most of the multicast traffic are carried in unicast "tunnels", where BGRP should treat such traffic transparently.

BGRP requires two independent phases to set up a reservation: path discovery and reservation aggregation. We describe these phases below.

7.1 Path Discovery

The reservation senders send PROBE messages to discover the reservation path. The PROBE messages traverse BR's hop-by-hop until reaching the destination domain. Each BGP router in the transit AS's must insert the associated AS number (and its own IP address) in the PROBE messages, but doesn't need to keep any state information.

PROBE messages must travel the path that is used for actual reservation. The PROBE message forwarding decision is made base on bi-lateral agreement between ISP's and some policy constraints at each BGP hop. The BGP hop can be found from the BGP Next-Hop attributes at each border.

The border routers send rejection messages back to the senders if the

destination is unreachable or reservation loops are detected.

7.2 Reservation Aggregation

A BR at the destination domain may receive PROBE messages from multiple reservation sources. It uses the information carried in the PROBE messages to construct an AS-level graph from which a loop-free sink tree is formed.

The BR sends reservation request (GRAFT) messages upstream toward the reservation sender. GRAFT messages traverse exactly the reversed list of BR's in the PROBE messages. Upon reception, transit routers can interface with intra-domain traffic-engineering protocols (such as RSVP, RSVP-LSP, or LDP) to set up a reservation within their AS's.

The transit BR's aggregate the reservations going to the same reservation receiver. The operation of reservation aggregation is illustrated in Section 4.

7.3 Reservation Management

Each BGRP-enabled router periodically sends REFRESH messages to its adjacent BGRP routers. If a particular reservation has not been refreshed within a period of time, it will be deleted and the associated resource will be freed. Each REFRESH message must contains a list of all reservation states at a BR. Each reservation state is compressed. A similar mechanism has been proposed for RSVP [25].

In case of route changes, BGP must up-call BGRP to re-adjust reservations. To reduce the effect of routing flapping, it may require some dampening mechanism to be applied on BGRP routers during reservation adjustment.

7.4 BGRP Enhancements

BGRP sends at least one PROBE message and one GRAFT message between leaf and root for each new reservation. Since these messages consume processing CPU and bandwidth, one would like to reducing the control message volume, and thereby add another dimension of scalability to BGRP. This can be done by making the following enhancements to the protocol.

CIDR labeling: Sink trees can be labeled with the CIDR prefixes associated with the tree root. An advancing PROBE message can recognize when it has reached the reservation tree it would like to join, and no need to propagate further.

Over-reservation: Sink tree nodes can reserve more bandwidth

between themselves and the tree root than is currently required. (One can think of this as aggregated advance reservations on behalf of unknown parties.) Each router keeps track of the actual reserved resources and the reservation it made downstream. It only forwards the PROBE, with a step increase in the reservation, if the actual committed resources reach the higher reservation level.

8 Relationship with Other Protocols

8.1 Comparing BGRP with RSVP

BGRPs approach is similar to RSVP, with BGRP's PROBE and GRAFT messages playing similar roles to RSVP's PATH and RESV messages. However, there are some of important differences between the two protocols:

RSVP PATH vs. BGRP PROBE: RSVP's PATH message primarily installs routing state at intermediate routers to guide RESV messages to the data senders. Routers must therefore keep both reservation sender and destination information. In a network with N nodes, this may require O(N2) entries. In BGRP, routers only store the reservation information for the O(N) sinks, but no source information.

Reservation aggregation: RSVP offers per-source and shared reservation styles. In the latter, multiple multicast senders take turns sharing a single reservation [26]. BGRP aggregates reservations by adding them together and propagating them downstream.

8.2 Interface with MPLS

RSVP and LDP are two protocols being designed to distribute MPLS labels and thus provide intra-domain level traffic engineering. BGRP interfaces with them at network border to trigger the establishment of MPLS LSP's.

8.3 Interface with End Users

We assume the ISP's set up multi-domain reservations in advance. End users use application-driven reservation protocols such as YESSIR [6] and RSVP [1] to request resource from the network. As the end-to-end reservation requests are received at the network border, the border routers simply aggregate the requests into the pre-computed trunks.

9 Example BGRP Scenarios

In this section, we illustrate the operation of BGRP with the network configuration in Figure-2.

9.1 Path Discovery

Initially, Net1 needs to reserve a path to Net6. The reservation can be characterized in terms of bandwidth, or a label path. We denote the reservation as R16. When the network gateway BR-A has been notified, it sends out a PROBE message with the following information:

```
Reservation Source:  BR-A
Network Destination: Net6
Reservation Amount:  R16
```

The PROBE message is encapsulated in TCP and delivered to an adjacent BGP next-hop router, BR-C.

Upon receiving the PROBE message, BR-C checks with the routing database and the bi-lateral agreements, and determines that the reservation should go through one of the BGP Next-Hops, BR-F. BR-C will insert its IP address into the PROBE and forward the PROBE to BR-F via TCP.

In turn, BR-F examines the PROBE against own resource database, insert its IP address, and forwards the PROBE to its BGP Next-Hop, BR-K.

At BR-K, the PROBE message looks like the following:

```
Reservation Source:  BR-A
Network Destination: Net6
Reservation Amount:  R16
Route Record:        {BR-C, BR-F}
```

After checking with own RIB (Routing Information Base), BR-I finds out that Net6 is an advertised IGP route, thus terminate the probing process.

9.2 Reservation Aggregation

BR-K can finish up the reservation process by sending a GRAFT message

upstream. The GRAFT message contains the following information.

```
Sink-Tree ID:        BR-K
Reservation Amount:  R16
Reservation Path:    {BR-F, BR-C, BR-A}
```

Before sending the GRAFT message, BR-K needs to reserve resource on the link to BR-F. The Reservation Path in GRAFT describes the explicit reservation path through the backbone. The GRAFT message is sent to the first reservation hop, BR-F.

On receiving the GRAFT message, BR-F first sets up an intra-domain reservation trunk between BR-F and the next reservation hop, BR-C. The ISP for Domain-3 can use any of the existing MPLS and resource reservation signaling protocols to setup such a trunk. After the trunk has been established, BR-F forwards the GRAFT to BR-C and stores the following information:

Tree-ID	Total Resv	Next-Hop	Previous-Hop	Branch Resv
BR-K	R16	BR-K	BR-C	R16

Similarly, BR-C sets up a reservation to BR-A. After getting the GRAFT message, BR-A terminates the reservation and direct all Net1 to Net6 user traffic through the new inter-domain reservation trunk.

Here let's assume that BR-B had probed the network for a reservation, R36, to Net6 on behalf of Net3. The final probed path is (BR-K, BR-F, BR-D, BF-B). BR-K sends a GRAFT message with the following information:

```
Sink-Tree ID:        BR-K
Reservation Amount:  R36
Reservation Path:    {BR-F, BR-D, BR-B}
```

Since this reservation shares the same sink-tree id as the previously described reservation, the reservations will aggregate at BR-F. The

end result is stored at the router, BR-F:

Tree-ID	Total Resv	Next-Hop	Previous-Hop	Branch Resv
BR-K	[R16 + R36]	BR-K	BR-C	R16
			BR-D	R36

The example shows the routers need to only store sink-tree information. Given that the average AS length in the Internet is somewhere between 4 and 5 [27], the storage and process gain introduced here can be very significant [13].

9.3 State Maintenance

BGRP uses "soft-state" to manage reservations. BGRP routers must send refresh messages to next-hop and previous hop routers periodically. When no refresh messages are received from a peer for a period of time, the BGRP router will delete the reserved resource.

In the example, if the BR-C to BR-F is broken, BR-C will stop sending refreshes to BR-F. Some time later, BR-F will adjust the reservation and update the local router information:

Tree-ID	Total Resv	Next-Hop	Previous-Hop	Branch Resv
BR-K	R36	BR-K	BR-D	R36

BR-F will propagate the reduced reservation information downstream in its next refresh cycle.

10 Security Considerations

In the BGRP model, we always assume some level of trust between BGRP routers. The reservation information is delivered domain by domain. Without proper authentication, this will enable denial of service attacks. Integrity information is required for each BGRP message.

11 Acknowledgments

Sean McCreary of NLANR/CAIDA helped us to collect and analyze the AS traces. Craig Labovitz advised us on various aspects of network operations that are relevant to BGRP. Tony Przygienda and Rohit Dube commented on the interaction between BGP and routing aggregation. We also thank Fred Baker, George Swallow, Roch Guerin, Geoff Huston and Andreas Terzis for discussions leading to this work.

12 Bibliography

[1] R. Braden, Ed., L. Zhang, S. Berson, S. Herzog, and S. Jamin, "Resource ReSerVation protocol (RSVP) -- version 1 functional specification," Request for Comments (Proposed Standard) 2205, Internet Engineering Task Force, Sept. 1997.

[2] D. Awduche, L. Berger, D. Gan, T. Li, G. Swallow, and V. Srinivasan, "Extensions to RSVP for LSP tunnels," Internet Draft, Internet Engineering Task Force, Mar. 1999. Work in progress.

[3] R. Callon, N. Feldman, A. Fredette, G. Swallow, and A. Viswanathan, "A framework for multiprotocol label switching," Internet Draft, Internet Engineering Task Force, June 1999. Work in progress.

[4] R. Callon, A. Viswanathan, and E. Rosen, "Multiprotocol label switching architecture," Internet Draft, Internet Engineering Task Force, Apr. 1999. Work in progress.

[5] D. Awduche, Y. Rekhter, J. Drake, and R. Coltun, "Multi-protocol lambda switching: Combining MPLS traffic engineering control with optical crossconnects," Internet Draft, Internet Engineering Task Force, Nov. 1999. Work in progress.

[6] P. Pan and H. Schulzrinne, "YESSIR: a simple reservation mechanism for the Internet," ACM Computer Communication Review, vol. 29, pp. 89--101, Apr. 1999.

[7] L. Zhang, S. Deering, D. Estrin, S. Shenker, and D. Zappala, "RSVP: a new resource reservation protocol," in Proceedings of the International Networking Conference (INET) , (San Francisco, California), pp. BCB--1, Internet Society, Aug. 1993.

[8] Internet Software Consortium, "Internet domain survey." http://www.isc.org/ds/.

[9] NLANR, "Nlanr network traffic packet header traces." http://moat.nlanr.net/Traces/.

[10] C. Labovitz, G. R. Malan, and F. Jahanian, "Internet routing

instability," in SIGCOMM Symposium on Communications Architectures and Protocols , (Cannes, France), Sept. 1997.

[11] T. Bates, "The cidr report." http://www.employees.org/~tbates/cidr-report.html.

[12] T. Li and Y. Rekhter, "A border gateway protocol 4 (BGP-4)," Internet Draft, Internet Engineering Task Force, Aug. 1998. Work in progress.

[13] P. Pan, E. Hahne, and H. Schulzrinne, "BGRP: A tree-based aggregation protocol for inter-domain reservations," Technical Report CUCS-029-99, Columbia University, New York, New York, Dec. 1999. http://www.cs.columbia.edu/~library/TR-repository/reports/reports-1999/cucs-029-99.pdf.

[14] J. Agogbua, D. Awduche, J. Malcolm, J. McManus, and M. O'Dell, "Requirements for traffic engineering over MPLS," Internet Draft, Internet Engineering Task Force, June 1999. Work in progress.

[15] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, "An architecture for differentiated service," Request for Comments (Informational) 2475, Internet Engineering Task Force, Dec. 1998.

[16] O. Schelen and S. Pink, "Aggregating resource reservation over multiple routing domains," in Proc. of Fifth IFIP International Workshop on Quality of Service (IwQOS) , (Cambridge, England), June 1998.

[17] S. Berson and S. Vincent, "Aggregation of internet integrated services state," in submitted to IWQOS , 1998.

[18] F. Reichmeyer, L. Ong, A. Terzis, L. Zhang, and R. Yavatkar, "A two-tier resource management model for differentiated services networks," Internet Draft, Internet Engineering Task Force, Nov. 1998. Work in progress.

[19] R. Guerin, S. Herzog, and S. Blake, "Aggregating RSVP-based QoS requests," Internet Draft, Internet Engineering Task Force, Nov. 1997. Work in progress.

[20] F. Baker, "Aggregation of RSVP for IP4 and IP6 reservations," Internet Draft, Internet Engineering Task Force, June 1999. Work in progress.

[21] G. Feher, K. Nemeth, M. Maliosz, I. Cselenyi, J. Bergkvist, D. Ahlard, and T. Engborg, "Boomerang - a simple protocol for resource

reservation in ip networks," in IEEE Workshop on QoS Support for Real-Time Internet Applications , (Vancouver, Canada), June 1999.

[22] Sprint, "Sprintlink policies." http://www.sprintlink.net/policies.htm.

[23] B. Suter and P. Pan, "COPS extension for intra-domain traffic engineering," Internet Draft, Internet Engineering Task Force, June 1999. Work in progress.

[24] N. Feldman, P. Doolan, L. Andersson, and A. Fredette, "LDP specification," Internet Draft, Internet Engineering Task Force, Dec. 1997. Work in progress.

[25] L. Berger, D. Gan, G. Swallow, and P. Pan, "RSVP refresh reduction extensions," Internet Draft, Internet Engineering Task Force, July 1999. Work in progress.

[26] D. J. Mitzel and S. Shenker, "Asymptotic resource consumption in multicast reservation styles," in SIGCOMM Symposium on Communications Architectures and Protocols , (London, UK), pp. 226--233, Sept. 1994.

[27] NLANR, "Nlanr as path lengths." http://moat.nlanr.net/ASPL/.

13 Authors

Ping Pan
Bell Labs, Lucent
101 Crawfords Corner Road
Holmdel, NJ 07733
USA
electronic mail: pingpan@research.bell-labs.com

Ellen Hahne
Bell Labs, Lucent
600 Mountain Ave.
Murray Hill, NJ 07974
USA
electronic mail: hahne@bell-labs.com

Henning Schulzrinne
Dept. of Computer Science
Columbia University
1214 Amsterdam Avenue
New York, NY 10027
USA

electronic mail: schulzrinne@cs.columbia.edu

Table of Contents

H:.tex