S. Berson
ISI
L. Berger
FORE Systems
March 1997

IP Integrated Services with RSVP over ATM

March 26, 1997

Status of Memo

Abstract

This draft describes a method for providing IP Integrated Services with RSVP over ATM switched virtual circuits (SVCs). It provides an overall approach to the problem as well as a specific method for running over today's ATM networks. There are two parts of this problem. This draft provides guidelines for using ATM VCs with QoS as part of an Integrated Services Internet. A related draft[6] describes service mappings between IP Integrated Services and ATM services.

Authors' Note

The postscript version of this document contains figures that are not included in the text version, so it is best to use the postscript version. Figures will be converted to ASCII in a future version.

Contents

1 Introduction

The Internet currently has one class of service normally referred to as "best effort." This service is typified by first-come, first-serve scheduling at each hop in the network. Best effort service has worked well for electronic mail, World Wide Web (WWW) access, file transfer (e.g. ftp), etc. For real-time traffic such as voice and video, the current Internet has performed well only across unloaded portions of the network. In order to provide guaranteed quality real-time traffic, new classes of service and a QoS signalling protocol are being introduced in the Internet[7, 18, 17], while retaining the existing best effort service. The QoS signalling protocol is RSVP[8, 19], the Resource ReSerVation Protocol.

ATM is rapidly becoming an important link layer technology. One of the important features of ATM technology is the ability to request a point-to-point Virtual Circuit (VC) with a specified Quality of Service (QoS). An additional feature of ATM technology is the ability to request point-to-multipoint VCs with a specified QoS. Point-to-multipoint VCs allows leaf nodes to be added and removed from the VC dynamically and so provide a mechanism for supporting IP multicast. It is only natural that RSVP and the Internet Integrated Services (IIS) model would like to utilize the QoS properties of any underlying link layer including ATM.

Classical IP over ATM[11] has solved part of this problem, supporting IP unicast best effort traffic over ATM. Classical IP over ATM is based on a Logical IP Subnetwork (LIS), which is a separately administered IP sub-network. Hosts within a LIS communicate using the ATM network, while hosts from different sub-nets communicate only by going through an IP router (even though it may be possible to open a direct VC between the two hosts over the ATM network). Classical IP over ATM provides an Address Resolution Protocol (ATMARP) for ATM edge devices to resolve IP addresses to native ATM addresses. For any pair of IP/ATM edge devices (i.e. hosts or routers), a single VC is created on demand and shared for all traffic between the two devices. A second part of the RSVP and IIS over ATM problem, IP multicast, is close to being solved with MARS[1], the Multicast Address Resolution Server. MARS compliments ATMARP by allowing an IP address to resolve into a list of native ATM addresses, rather than just a single address.

A key remaining issue for IP over ATM is the integration of RSVP signalling and ATM signalling in support of the Internet Integrated Services (IIS) model. There are two main areas involved in supporting the IIS model, QoS translation and VC management. QoS translation concerns mapping a QoS from the IIS model to a proper ATM QoS, while VC management concentrates on how many VCs are needed and which traffic flows are routed over which VCs. Mapping of IP QoS to ATM QoS is the subject of a companion draft[6].

This draft concentrates on VC management (and we assume in this draft that the QoS for a single reserved flow can be acceptably translated to an ATM QoS). Two types of VCs need to be managed, data VCs which handle the actual data traffic, and control VCs which handle the RSVP signalling traffic. Several VC management schemes for both data and control VCs are described in this draft. For each scheme, there are two major issues - (1) heterogeneity and (2) dynamic behavior. Heterogeneity refers to how requests for different QoS's are handled, while dynamic behavior refers to how changes in QoS and changes in multicast group membership are handled. These schemes will be evaluated in terms of the following metrics - (1) number of VCs needed to implement the scheme, (2) bandwidth wasted due to duplicate packets, and (3) flexibility in handling heterogeneity and dynamic behavior. Examples where each scheme is most applicable are given.

1.1 Terms

The terms "reservation" and "flow" are used in many contexts, often with different meaning. These terms are used in this document with the following meaning:

- **Reservation** is used in this document to refer to an RSVP initiated request for resources. RSVP initiates requests for resources based on RESV message processing. RESV messages that simply refresh state do not trigger resource requests. Resource requests may be made based on RSVP sessions and RSVP reservation styles. RSVP styles dictate whether the reserved resources are used by one sender or shared by multiple senders. See [8] for details of each. Each new request is referred to in this document as an RSVP reservation, or simply reservation.
- **Flow** is used to refer to the data traffic associated with a particular reservation. The specific meaning of flow is RSVP style dependent. For shared style reservations, there is one flow per session. For distinct style reservations, there is one flow per sender (per session).

1.2 Assumptions

The following assumptions are made:

- Support for IPv4 and IPv6 best effort in addition to QoS
- Use RSVP with policy control as signalling protocol
- Assume UNI 3.x and 4.0 ATM services
- VCs initiation by sub-net senders

1.2.1 IPv4 and IPv6

Currently IPv4 is the standard protocol of the Internet which now provides only best effort service. We assume that best effort service will continue to be supported while introducing new types of service according to the IP Integrated Services model. We also assume that IPv6 will be supported as well as IPv4.

1.2.2 RSVP and Policy

We assume RSVP as the Internet signalling protocol which is described in [19]. The reader is assumed to be familiar with [19].

IP Integrated Services discriminates between users by providing some users better service at the expense of others. For ATM, preferential service reflects when a new VC is created for a user rather than joining an existing VC. Policy determines how preferential services are allocated while allowing network operators maximum flexibility to provide value-added services for the marketplace. Mechanisms need to be be provided to enforce access policies. These mechanisms may include such things as permissions and/or billing.

For scaling reasons, policies based on bilateral agreements between neighboring providers are considered. The bilateral model has similar scaling properties to multicast while maintaining no global information. Policy control is currently being developed for RSVP (see [10] for details).

1.2.3 ATM

We assume ATM defined by UNI 3.x and 4.0, plus TM 4.0. ATM provides both point-to-point and point-to-multipoint Virtual Circuits (VCs) with a specified Quality of Service (QoS). ATM provides both Permanent Virtual Circuits (PVCs) and Switched Virtual Circuits (SVCs). In the Permanent Virtual Circuit (PVC) environment, PVCs are typically used as point-to-point link replacements. So the Integrated Services support issues are similar to point-to-point links. This draft describes schemes for supporting Integrated Services using SVCs.

1.2.4 VC Initiation

There is an apparent mismatch between RSVP and ATM. Specifically, RSVP control is receiver oriented and ATM control is sender oriented. This initially may seem like a major issue, but really is not. While RSVP reservation (RESV) requests are generated at the receiver, actual allocation of resources takes place at the sub-net sender.

For data flows, this means that sub-net senders will establish all QoS VCs and the sub-net receiver must be able to accept incoming QoS VCs. These restrictions are consistent with RSVP version 1 processing rules and allow senders to use different flow to VC mappings and even different QoS renegotiation techniques without interoperability problems. All RSVP over ATM approaches that have VCs initiated and controlled by the sub-net senders will interoperate. Figure 1 shows this model of data flow VC initiation.

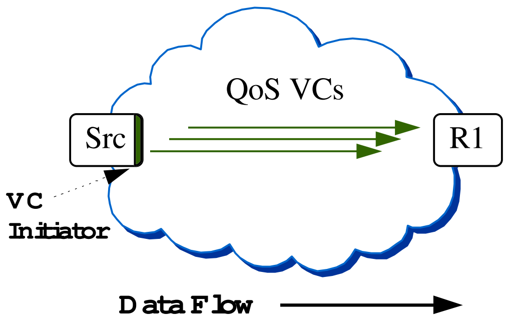

Figure 1: Data Flow VC Initiation

The use of the reverse path provided by point-to-point VCs by receivers is for further study. There are two related issues. The first is that use of the reverse path requires the VC initiator to set appropriate reverse path QoS parameters. The second issue is that reverse paths are not available with point-to-multipoint VCs, so reverse paths could only be used to support unicast RSVP reservations. Receivers initiating VCs via the reverse path mechanism provided by point-to-point VCs is also for future study.

2 Policy

RSVP allows for local policy control [10] as well as admission control. Thus a user can request a reservation with a specific QoS and with a policy object that, for example, offers to pay for additional costs setting up a new reservation. The policy module at the entry to a provider can decide how to satisfy that request - either by merging the request in with an existing reservation or by creating a new reservation for this (and perhaps other) users. This policy can be on a per user-provider basis where a user and a provider have an agreement on the type of service offered, or on a provider-provider basis, where two providers have such an agreement. With the ability to do local policy control, providers can offer services best suited to their own resources and their customers needs.

Policy is expected to be provided as a generic API which will return values indicating what action should be taken for a specific reservation request. The API is expected to have access to the reservation tables with the QoS for each reservation. The RSVP Policy and Integrity objects will be passed to the policy() call. Four possible return values are expected. The request can be rejected. The request can be accepted as is. The request can be accepted but at a different QoS. The request can cause a change of QoS of an existing reservation. The information returned from this call will be used to call the admission control interface. As this area is critical to deployment, progress will need to be made in this area.

3 Data VC Management

Any RSVP over ATM implementation must map RSVP and RSVP associated data flows to ATM Virtual Circuits (VCs). LAN Emulation [2], Classical IP [11] and, more recently, NHRP [12] discuss mapping IP traffic onto ATM SVCs, but they only cover a single QoS class, i.e., best effort traffic. When QoS is introduced, VC mapping must be revisited. For RSVP controlled QoS flows, one issue is VCs to use for QoS data flows.

In the Classic IP over ATM and current NHRP models, a single point-to-point VC is used for all traffic between two ATM attached hosts (routers and end-stations). It is likely that such a single VC will not be adequate or optimal when supporting data flows with multiple QoS types. RSVP's basic purpose is to install support for flows with multiple QoS types, so it is essential for any RSVP over ATM solution to address VC usage for QoS data flows.

This section describes issues and methods for management of VCs associated with QoS data flows. When establishing and maintaining VCs, the sub-net sender will need to deal with several complicating factors including multiple QoS reservations, requests for QoS changes, ATM short-cuts, and several multicast specific issues. The multicast specific issues result from the nature of ATM connections. The key multicast related issues are heterogeneity, data distribution, receiver transitions, and end-point identification.

3.1 Reservation to VC Mapping

There are various approaches available for mapping reservations on to VCs. A distinguishing attribute of all approaches is how reservations are combined on to individual VCs. When mapping reservations on to VCs, individual VCs can be used to support a single reservation, or reservation can be combined with others on to "aggregate" VCs. In the first case, each reservation will be supported by one or more VCs. Multicast reservation requests may translate into the setup of multiple VCs as is described in more detail in section 3.2. Unicast reservation requests will always translate into the setup of a single QoS VC. In both cases, each VC will only carry data associated with a single reservation. The greatest benefit if this approach is ease of implementation, but it comes at the cost of increased (VC) setup time and the consumption of greater number of VC and associated resources.

We refer to the other case, when reservations are not combined, as the "aggregation" model. With this model, large VCs could be set up between IP routers and hosts in an ATM network. These VCs could be managed much like IP Integrated Service (IIS) point-to-point links (e.g. T-1, DS-3) are managed now. Traffic from multiple sources over multiple RSVP sessions might be multiplexed on the same VC. This approach has a number of advantages. First, there is typically no signalling latency as VCs would be in existence when the traffic started flowing, so no time is wasted in setting up VCs. Second, the heterogeneity problem (section 3.2) in full over ATM has been reduced to a solved problem. Finally, the dynamic QoS problem (section 3.6) for ATM has also been reduced to a solved problem. This approach can be used with point-to-point and point-to-multipoint VCs. The problem with the aggregation approach is that the choice of what QoS to use for which of the VCs is difficult, but is made easier since the VCs can be changed as needed. The advantages of this scheme makes this approach an item for high priority study.

3.2 Heterogeneity

Heterogeneity occurs when receivers request different QoS's within a single session. This means that the amount of requested resources differs on a per next hop basis. A related type of heterogeneity occurs due to best-effort receivers. In any IP multicast group, it is possible that some receivers will request QoS (via RSVP) and some receivers will not. Both types of heterogeneity are shown in figure 2. In shared media, like Ethernet, receivers that have not requested resources can typically be given identical service to those that have without complications. This is not the case with ATM. In ATM networks, any additional end-points of a VC must be explicitly added. There may be costs associated with adding the best-effort receiver, and there might not be adequate resources. An RSVP over ATM solution will need to support heterogeneous receivers even though ATM does not currently provide such support directly.

RSVP heterogeneity is supported over ATM in the way RSVP reservations are mapped into ATM VCs. There are multiple models for supporting RSVP heterogeneity over ATM. Section 3.2.1 examines the multiple VCs per RSVP reservation (or full heterogeneity) model where a single reservation can be forwarded into several VCs each with a different QoS. Section 3.2.2 presents a limited heterogeneity model where exactly one QoS VC is used along with a best effort VC. Section 3.2.3 examines the VC per RSVP reservation (or homogeneous) model, where each RSVP reservation is mapped to a single ATM VC. Section 3.2.4 describes the aggregation model allowing

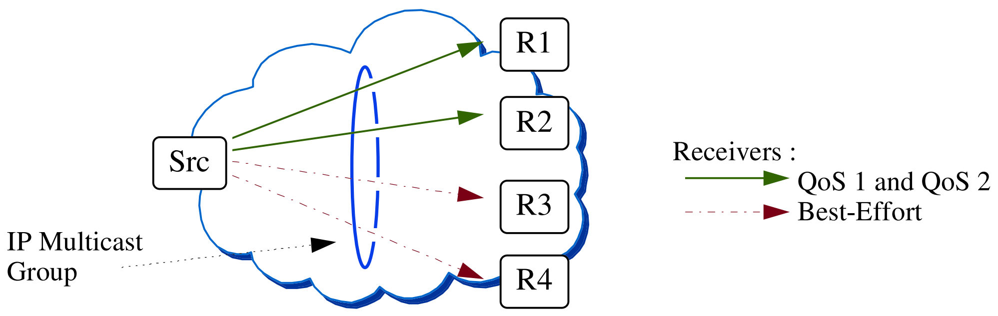

Figure 2: Types of Multicast Receivers

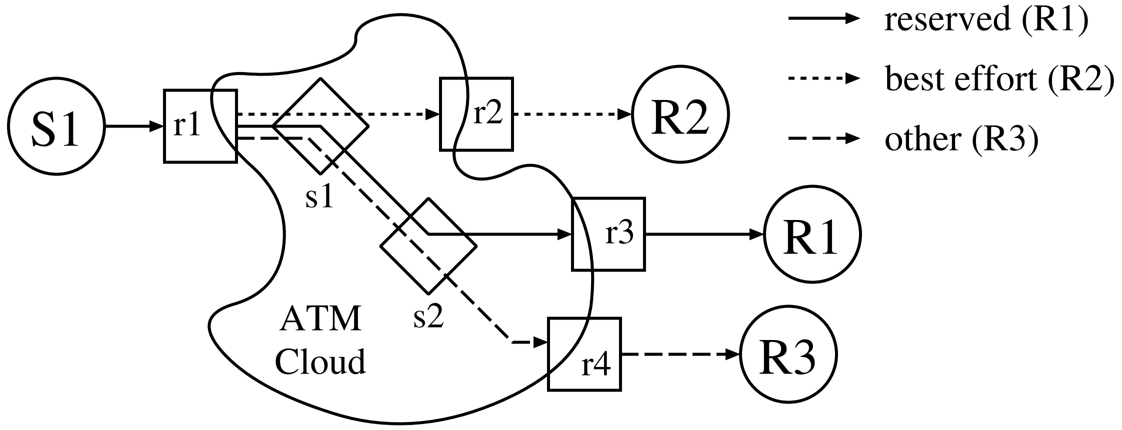

Figure 3: Full heterogeneity

aggregation of multiple RSVP reservations into a single VC. Further study is being done on the aggregation model.

3.2.1 Full Heterogeneity Model

We define the "full heterogeneity" model as providing a separate VC for each distinct QoS for a multicast session including best effort and one or more QoS's. This is shown in figure 3 where S1 is a sender, R1-R3 are receivers, r1-r4 are IP routers, and s1-s2 are ATM switches. Receivers R1 and R3 make reservations with different QoS while R2 is a best effort receiver. Three point-to-multipoint VCs are created for this situation, each with the requested QoS. Note that any leafs requesting QoS 1 or QoS 2 would be added to the existing QoS VC.

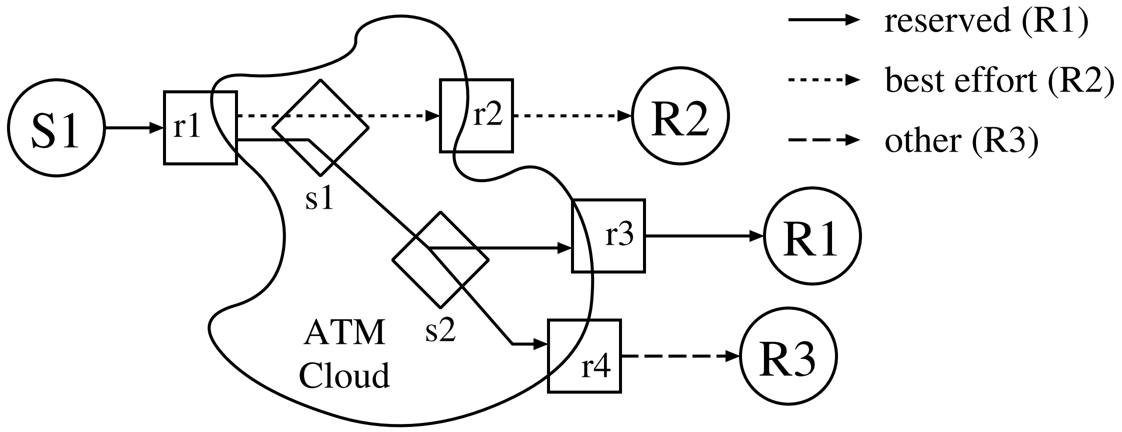

Figure 4: Limited heterogeneity

Note that while full heterogeneity gives users exactly what they request, it requires more resources of the network than other possible approaches. In figure 3, three copies of each packet are sent on the link from r1 to s1. Two copies of each packet are then sent from s1 to s2. The exact amount of bandwidth used for duplicate traffic depends on the network topology and group membership.

3.2.2 Limited Heterogeneity Model

An important special case of full heterogeneity is limited heterogeneity. We define the "limited heterogeneity" model as the case where the receivers of a multicast session are limited to use either best effort service or a single alternate quality of service. The alternate QoS can be chosen either by higher level protocols or by dynamic renegotiation of QoS as described below.

In order to support limited heterogeneity, each ATM edge device participating in a session would need at most two VCs. One VC would be a point-to-multipoint best effort service VC and would serve all best effort service IP destinations for this RSVP session. The other VC would be a point to multipoint VC with QoS and would serve all IP destinations for this RSVP session that have an RSVP reservation established. This is shown in figure 4 where there are three receivers, R2 requesting best effort service, while R1 and R3 request distinct reservations. Whereas, in figure 3, R1 and R3 have a separate VC, so each receives precisely the resources requested, in figure 4, R1 and R3 share the same VC (using the maximum of R1 and R3 QoS) across the ATM network. Note that though the VC and hence the QoS for R1 and R3 are the same within the ATM cloud, the reservation outside the ATM cloud (from router r4 to receiver R3) uses the QoS actually requested by R3.

As with full heterogeneity, a disadvantage of the limited heterogeneity scheme is that each packet will need to be duplicated at the network layer and one copy sent into each of the 2 VCs. Again, the exact amount of excess traffic will depend on the network topology and group membership. Looking at figure 4, there are two VCs going from router r1 to switch s1. Two copies of every packet will traverse the r1-s1 link. Another disadvantage of limited heterogeneity is that a reservation request can be rejected even when the resources are available. This occurs when a new receiver requests a larger QoS. If any of the existing QoS VC end-points cannot upgrade to the new QoS, then the new reservation fails though the resources exist for the new receiver.

3.2.3 Homogeneous and Modified Homogeneous Models

We define the "homogeneous" model as the case where all receivers of a multicast session use a single quality of service VC. Best-effort receivers also use the single RSVP triggered QoS VC. The single VC can be a point-to-point or point-to-multipoint as appropriate. The QoS VC is sized to provide the maximum resources requested by all RSVP next-hops.

This model matches the way the current RSVP specification addresses heterogeneous requests. The current processing rules and traffic control interface describe a model where the largest requested reservation for a specific outgoing interface is used in resource allocation, and traffic is transmitted at the higher rate to all next-hops. This approach would be the simplest method for RSVP over ATM implementations.

While this approach is simple to implement, providing better than best-effort service may actually be the opposite of what the user desires since in providing ATM QoS. There may be charges incurred or resources that are wrongfully allocated. There are two specific problems. The first problem is that a user making a small or no reservation would share a QoS VC resources without making (and perhaps paying for) an RSVP reservation. The second problem is that a receiver may not receive any data. This may occur when there is insufficient resources to add a receiver. The rejected user would not be added to the single VC and it would not even receive traffic on a best effort basis.

Not sending data traffic to best-effort receivers because of another receiver's RSVP request is clearly unacceptable. The previously described limited heterogeneous model ensures that data is always sent to both QoS and best-effort receivers, but it does so by requiring replication of data at the sender in all cases. It is possible to extend the homogeneous model to both ensure that data is always sent to best-effort receivers and also to avoid replication in the normal case. This extension is to add special handling for the case where a best-effort receiver cannot be added to the QoS VC. In this case, a best-effort VC can be established to any receivers that could not be added to the QoS VC. Only in this special error case would senders be required to replicate data. We define this approach as the "modified homogeneous" model.

3.2.4 Aggregation

The last scheme is the multiple RSVP reservations per VC (or aggregation) model. With this model, large VCs could be set up between IP routers and hosts in an ATM network. These VCs could be managed much like IP Integrated Service (IIS) point-to-point links (e.g. T-1, DS-3) are managed now. Traffic from multiple sources over multiple RSVP sessions might be multiplexed on the same VC. This approach has a number of advantages. First, there is typically no signalling latency as VCs would be in existence when the traffic started flowing, so no time is wasted in setting up VCs. Second, the heterogeneity problem in full over ATM has been reduced to a solved problem. Finally, the dynamic QoS problem for ATM has also been reduced to a solved problem. This approach can be used with point-to-point and point-to-multipoint VCs. The problem with the aggregation approach is that the choice of what QoS to use for which of the VCs is difficult, but is made easier since the VCs can be changed as needed. The advantages of this scheme makes this approach an item for high priority study.

Multiple options for mapping reservations onto VCs have been discussed. No matter which model or combination of models is used by an implementation, implementations must not normally send more than one copy of a particular data packet to a particular next-hop (ATM end-point). Some transient over transmission is acceptable, but only during VC setup and transition. Implementations must also ensure that data traffic is sent to best-effort receivers. Data traffic may be sent to best-effort receivers via best-effort or QoS VCs as is appropriate for the implemented model. In all cases, implementations must not create VCs in such a way that data cannot be sent to best-effort receivers. This includes the case of not being able to add a best-effort receiver to a QoS VC, but does not include the case where best-effort VCs cannot be setup. The failure to establish best-effort VCs is considered to be a general IP over ATM failure and is therefore beyond the scope of this document.

3.3 Multicast End-Point Identification

Implementations must be able to identify ATM end-points participating in an IP multicast group. The ATM end-points will be IP multicast receivers and/or next-hops. Both QoS and best-effort end-points must be identified. RSVP next-hop information will provide QoS end-points, but not best-effort end-points.

Another issue is identifying end-points of multicast traffic handled by non-RSVP capable next-hops. In this case a PATH message travels through a non-RSVP egress router on the way to the next hop RSVP node. When the next hop RSVP node sends a RESV message it may arrive at the source over a different route than what the data is using. The source will get the RESV message, but will not know which egress router needs the QoS. For unicast sessions, there is no problem since the

ATM end-point will be the IP next-hop router. Unfortunately, multicast routing may not be able to uniquely identify the IP next-hop router. So it is possible that a multicast end-point can not be identified.

In the most common case, MARS will be used to identify all end-points of a multicast group. In the router to router case, a multicast routing protocol may provide all next-hops for a particular multicast group. In either case, RSVP over ATM implementations must obtain a full list of end-points, both QoS and non-QoS, using the appropriate mechanisms. The full list can be compared against the RSVP identified end-points to determine the list of best-effort receivers.

There is no straightforward solution to uniquely identifying end-points of multicast traffic handled by non-RSVP next hops. The preferred solution is to use multicast routing protocols that support unique end-point identification. In cases where such routing protocols are unavailable, all IP routers that will be used to support RSVP over ATM should support RSVP.

3.4 Multicast Data Distribution

Two models are planned for IP multicast data distribution over ATM. In one model, senders establish point-to-multipoint VCs to all ATM attached destinations, and data is then sent over these VCs. This model is often called “multicast mesh” or “VC mesh” mode distribution. In the second model, senders send data over point-to-point VCs to a central point and the central point relays the data onto point-to-multipoint VCs that have been established to all receivers of the IP multicast group. This model is often referred to as “multicast server” mode distribution. Figure 5 shows data flow for both modes of IP multicast data distribution. RSVP over ATM solutions must ensure that IP multicast data is distributed with appropriate QoS.

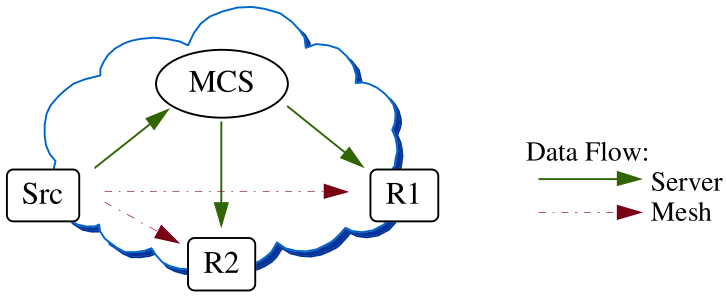

Figure 5: IP Multicast Data Distribution Over ATM

In the Classical IP context, multicast server support is provided via MARS[1]. MARS does not currently provide a way to communicate QoS requirements to a MARS multicast server (MCS).

When using multicast servers that do not support QoS requests, a sender must set the service, not global, break bit(s).

3.5 Receiver Transitions

When setting up a point-to-multipoint VCs there will be a time when some receivers have been added to a QoS VC and some have not. During such transition times it is possible to start sending data on the newly established VC. The issue is when to start send data on the new VC. If data is sent both on the new VC and the old VC, then data will be delivered with proper QoS to some receivers and with the old QoS to all receivers. This means the QoS receivers would get duplicate data. If data is sent just on the new QoS VC, the receivers that have not yet been added will lose information. So, the issue comes down to whether to send to both the old and new VCs, or to send to just one of the VCs. In one case duplicate information will be received, in the other some information may not be received. This issue needs to be considered for three cases: when establishing the first QoS VC, when establishing a VC to support a QoS change, and when adding a new end-point to an already established QoS VC.

The first two cases are very similar. It both, it is possible to send data on the partially completed new VC, and the issue of duplicate versus lost information is the same.

The last case is when an end-point must be added to an existing QoS VC. In this case the end-point must be both added to the QoS VC and dropped from a best-effort VC. The issue is which to do first. If the add is first requested, then the end-point may get duplicate information. If the drop is requested first, then the end-point may loose information.

In order to ensure predictable behavior and delivery of data to all receivers, data can only be sent on a new VCs once all parties have been added. This will ensure that all data is only delivered once to all receivers. This approach does not quite apply for the last case. In the last case, the add should be completed first, then the drop. This means that receivers must be prepared to receive some duplicate packets at times of QoS setup.

3.6 Dynamic QoS

RSVP provides dynamic quality of service (QoS) in that the resources that are requested may change at any time. There are several common reasons for a change of reservation QoS. First, an existing receiver can request a new larger (or smaller) QoS if the current received quality is unacceptable. Second, a sender may change its traffic specification (TSpec), which can trigger a change in the reservation requests of the receivers. Third, a new sender can start sending to a multicast group with a larger traffic specification than existing senders, triggering larger reservations. Finally, a new receiver can make a reservation that is larger than existing reservations. If the limited heterogeneity model is being used and the merge node for the larger reservation is an ATM edge device, a new larger reservation must be set up across the ATM network.

Since ATM service, as currently defined in UNI 3.x and UNI 4.0, does not allow renegotiating the QoS of a VC, dynamically changing the reservation means creating a new VC with the new QoS, and tearing down an established VC. Tearing down a VC and setting up a new VC in ATM are complex operations that involve a non-trivial amount of processor time, and may have a substantial latency.

There are several options for dealing with this mismatch in service. A specific approach will need to be a part of any RSVP over ATM solution.

One method for supporting changes in RSVP reservations is to attempt to replace an existing VC with a new appropriately sized VC. During setup of the replacement VC, the old VC must be left in place unmodified. The old VC is left unmodified to minimize interruption of QoS data delivery. Once the replacement VC is established, data transmission is shifted to the new VC, and the old VC is then closed.

If setup of the replacement VC fails, then the old QoS VC should continue to be used. When the new reservation is greater than the old reservation, the reservation request should be answered with an error. When the new reservation is less than the old reservation, the request should be treated as if the modification was successful. While leaving the larger allocation in place is suboptimal, it maximizes delivery of service to the user. Implementations should retry replacing the too large VC after some appropriate elapsed time.

One additional issue is that only one QoS change can be processed at one time per reservation. If the (RSVP) requested QoS is changed while the first replacement VC is still being setup, then the replacement VC is released and the whole VC replacement process is restarted.

To limit the number of changes and to avoid excessive signalling load, implementations may limit the number of changes that will be processed in a given period. One implementation approach would have each ATM edge device configured with a time parameter τ (which can change over time) that gives the minimum amount of time the edge device will wait between successive changes of the QoS of a particular VC. Thus if the QoS of a VC is changed at time t, all messages that would change the QoS of that VC that arrive before time $t+\tau$ would be queued. If several messages changing the QoS of a VC arrive during the interval, redundant messages can be discarded. At time $t+\tau$, the remaining change(s) of QoS, if any, can be executed.

The sequence of events for a single VC would be

1. Wait if timer is active
2. Establish VC with new QoS
3. Remap data traffic to new VC
4. Tear down old VC
5. Activate timer

There is an interesting interaction between heterogeneous reservations and dynamic QoS. In the case where a RESV message is received from a new next-hop and the requested resources are larger than any existing reservation, both dynamic QoS and heterogeneity need to be addressed. A key issue is whether to first add the new next-hop or to change to the new QoS. This is a fairly straight forward special case. Since the older, smaller reservation does not support the new next-hop, the dynamic QoS process should be initiated first. Since the new QoS is only needed by the new next-hop, it should be the first end-point of the new VC. This way signalling is minimized when the setup to the new next-hop fails.

3.7 Short-Cuts

Short-cuts [12] allow ATM attached routers and hosts to directly establish point-to-point VCs across LIS boundaries, i.e., the VC end-points are on different IP sub-nets. The ability for short-cuts and RSVP to interoperate has been raised as a general question. The area of concern is the ability to handle asymmetric short-cuts. Specifically how RSVP can handle the case where a downstream short-cut may not have a matching upstream short-cut. In this case, which is shown in figure 6, PATH and RESV messages following different paths.

Examination of RSVP shows that the protocol already includes mechanisms that will support short-cuts. The mechanism is the same one used to support RESV messages arriving at the wrong router and the wrong interface. The key aspect of this mechanism is RSVP only processing messages that arrive at the proper interface and RSVP forwarding of messages that arrive on the wrong interface. The proper interface is indicated in the NHOP object of the message. So, existing RSVP mechanisms will support asymmetric short-cuts.

The short-cut model of VC establishment still poses several issues when running with RSVP. The major issues are dealing with established best-effort short-cuts, when to establish short-cuts, and QoS only short-cuts. These issues will need to be addressed by RSVP implementations.

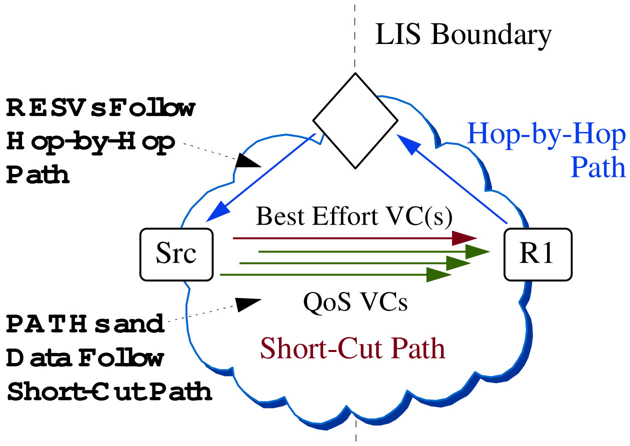

Figure 6: Asymmetric RSVP Message Forwarding With ATM Short-Cuts

3.8 VC Teardown

RSVP can identify from either explicit messages or timeouts when a data VC is no longer needed. Therefore, data VCs set up to support RSVP controlled flows should only be released at the direction of RSVP. VCs must not be timed out due to inactivity by either the VC initiator or the VC receiver. This conflicts with VCs timing out as described in RFC 1755[14], section 3.4 on VC Teardown. RFC 1755 recommends tearing down a VC that is inactive for a certain length of time. Twenty minutes is recommended. This timeout is typically implemented at both the VC initiator and the VC receiver. Although, section 3.1 of the update to RFC 1755[15] states that inactivity timers must not be used at the VC receiver.

When this timeout occurs for an RSVP initiated VC, a valid VC with QoS will be torn down unexpectedly. While this behavior is acceptable for best-effort traffic, it is important that RSVP controlled VCs not be torn down. If there is no choice about the VC being torn down, the RSVP daemon must be notified, so a reservation failure message can be sent.

4 RSVP Control VC Management

One last important issue is providing a data path for the RSVP messages themselves. There are two main types of messages in RSVP, PATH and RESV. PATH messages are sent to a multicast address, while RESV messages are sent to a unicast address. Other RSVP messages are handled similar to either PATH or RESV[1]. So ATM VCs used for RSVP signalling messages need to provide both unicast and multicast functionality.

There are several different approaches for how to assign VCs to use for RSVP signalling messages. The main approaches are:

- use same VC as data
- single VC per session
- single point-to-multipoint VC multiplexed among sessions
- multiple point-to-point VCs multiplexed among sessions

There are several different issues that affect the choice of how to assign VCs for RSVP signalling. One issue is the number of additional VCs needed for RSVP signalling. Related to this issue is the degree of multiplexing on the RSVP VCs. In general more multiplexing means less VCs. An additional issue is the latency in dynamically setting up new RSVP signalling VCs. A final issue is complexity of implementation. The remainder of this section discusses the issues and tradeoffs among these different approaches and suggests guidelines for when to use which alternative.

4.1 Mixed data and control traffic

In this scheme RSVP signalling messages are sent on the same VCs as is the data traffic. The main advantage of this scheme is that no additional VCs are needed beyond what is needed for the data traffic. An additional advantage is that there is no ATM signalling latency for PATH messages (which follow the same routing as the data messages). However there can be a major problem when data traffic on a VC is nonconforming. With nonconforming traffic, RSVP signalling messages may be dropped. While RSVP is resilient to a moderate level of dropped messages, excessive drops would lead to repeated tearing down and re-establishing QoS VCs, a very undesirable behavior for ATM. Due to these problems, this is not a good choice for providing RSVP signalling messages, even though the number of VCs needed for this scheme is minimized.

One variation of this scheme is to use the best effort data path for signalling traffic. In this scheme, there is no issue with nonconforming traffic, but there is an issue with congestion in the ATM network.

RSVP provides some resiliency to message loss due to congestion, but RSVP control messages should be offered a preferred class of service. A related variation of this scheme that is hopeful but

[1]This can be slightly more complicated for RERR messages

requires further study is to have a packet scheduling algorithm (before entering the ATM network) that gives priority to the RSVP signalling traffic. This can be difficult to do at the IP layer. One possible approach at the ATM layer would be to use the Cell Loss Priority (CLP) bit for RSVP signalling traffic to ensure better service.

4.2 Single RSVP VC per RSVP Reservation

In this scheme, there is a parallel RSVP signalling VC for each RSVP reservation. This scheme results in twice the minimum number of VCs, but means that RSVP signalling messages have the advantage of a separate VC. This separate VC means that RSVP signalling messages have their own traffic contract and compliant signalling messages are not subject to dropping due to other noncompliant traffic (such as can happen with the scheme in section 4.1). The advantage of this scheme is its simplicity - whenever a data VC is created, a separate RSVP signalling VC is created. The disadvantage of the extra VC is that extra ATM signalling needs to be done.

Additionally, this scheme requires twice the minimum number of VCs and also additional latency, but is quite simple. This approach would tend to work well on hosts.

4.3 Multiplexed point-to-multipoint RSVP VCs

In this scheme, there is a single point-to-multipoint RSVP signalling VC for each unique ingress router and unique set of egress routers. This scheme allows multiplexing of RSVP signalling traffic that shares the same ingress router and the same egress routers. This can save on the number of VCs, by multiplexing, but there are problems when the destinations of the multiplexed point-to-multipoint VCs are changing. Several alternatives exist in these cases, that have applicability in different situations. First, when the egress routers change, the ingress router can check if it already has a point-to-multipoint RSVP signalling VC for the new list of egress routers. If the RSVP signalling VC already exists, then the RSVP signalling traffic can be switched to this existing VC. If no such VC exists, one approach would be to create a new VC with the new list of egress routers. Other approaches include modifying the existing VC to add an egress router or using a separate new VC for the new egress routers. When a destination drops out of a group, an alternative would be to keep sending to the existing VC even though some traffic is wasted.

The number of VCs used in this scheme is a function of traffic patterns across the ATM network, but is always less than the number used with the Single RSVP VC per data VC. In addition, existing best effort data VCs could be used for RSVP signalling. Reusing best effort VCs saves on the number of VCs at the cost of higher probability of RSVP signalling packet loss. One possible place where this scheme will work well is in the core of the network where there is the most opportunity to take advantage of the savings due to multiplexing. The exact savings depend on the patterns of traffic and the topology of the ATM network.

4.4 Multiplexed point-to-point RSVP VCs

In this scheme, multiple point-to-point RSVP signalling VCs are used for a single point-to-multipoint data VC. This scheme allows multiplexing of RSVP signalling traffic but requires the same traffic to be sent on each of several VCs. This scheme is quite flexible and allows a large amount of multiplexing. Since point-to-point VCs can set up a reverse channel at the same time as setting up the forward channel, this scheme could save substantially on signalling cost. In addition, signalling traffic could share existing best effort VCs. Sharing existing best effort VCs reduces the total number of VCs needed, but might cause signalling traffic drops if there is congestion in the ATM network.

This point-to-point scheme would work well in the core of the network where there is much opportunity for multiplexing. Also in the core of the network, RSVP VCs can stay permanently established either as Permanent Virtual Circuits (PVCs) or as long lived Switched Virtual Circuits (SVCs). The number of VCs in this scheme will depend on traffic patterns, but in the core of a network would be approximately $n(n-1)/2$ where n is the number of IP nodes in the network. In the core of the network, this will typically be small compared to the total number of VCs.

4.5 QoS for RSVP VCs

There is an issue for what QoS, if any, to assign to the RSVP VCs. Three solutions have been covered in section 4.1 and in the shared best effort VC variations in sections 4.4 and 4.3. For other RSVP VC schemes, a QoS (possibly best effort) will be needed. What QoS to use partially depends on the expected level of multiplexing that is being done on the VCs, and the expected reliability of best effort VCs. Since RSVP signalling is infrequent (typically every 30 seconds), only a relatively small QoS should be needed. This is important since using a larger QoS risks the VC setup being rejected for lack of resources. Falling back to best effort when a QoS call is rejected is possible, but if the ATM net is congested, there will likely be problems with RSVP packet loss on the best effort VC also. Additional experimentation is needed in this area.

5 Encapsulation

Since RSVP is a signalling protocol used to control flows of IP data packets, encapsulation for both RSVP packets and associated IP data packets must be defined. There are currently two encapsulation options for running IP over ATM, RFC 1483 and LANE. There is also the possibility of future encapsulation options, such as MPOA[3]. The first option is described in RFC 1483[9] and is currently used for "Classical" IP over ATM and NHRP.

The second option is LAN Emulation, as described in [2]. LANE encapsulation does not currently include a QoS signalling interface. If LANE encapsulation is needed, LANE QoS signalling would first need to be defined by the ATM Forum. It is possible that LANE 2.0 will include the required QoS support.

6 Security

The same considerations stated in [8] and [14] apply to this document. There are no additional security issues raised in this document.

7 Future Work

We have described a set of schemes for deploying RSVP over IP over ATM. There are a number of other issues that are subjects of continuing research. These issues (and others) are covered in [5], and are briefly repeated here.

A major issue is providing policy control for ATM VC creation. There is work going on in the RSVP working group [8] on defining an architecture for policy support. Further work is needed in defining an API and policy objects. As this area is critical to deployment, progress will need to be made in this area.

NHRP provides advantages in allowing short-cuts across 2 or more LIS's. Short cutting router hops can lead to more efficient data delivery. Work on NHRP is on-going, but currently provides only a unicast delivery service. Further study is needed to determine how NHRP can be used with RSVP and ATM. Future work depends on the development of NHRP for multicast.

Furthermore, when using RSVP it may be desirable to establish multiple short-cut VCs, to use these VCs for specific QoS flows, and to use the hop-by-hop path for other QoS and non-QoS flows. The current NHRP specification [12] does not preclude such an approach, but nor does it explicitly support it. We believe that explicit support of flow based short-cuts would improve RSVP over ATM solutions. We also believe that such support may require the ability to include flow information in the NHRP request.

There is work in the ION working group on MultiCast Server (MCS) architectures for MARS. An MCS provides savings in the number of VCs in certain situations. When using a multicast server, the sub-network sender could establish a point-to-point VC with a specific QoS to the server, but there is not current mechanism to relay QoS requirements to the MCS. Future work includes providing RSVP and ATM support over MARS MCS's.

Unicast ATM VCs are inherently bi-directional and have the capability of supporting a *reverse channel*. By using the reverse channel for unicast VCs, the number of VCs used can potentially be reduced. Future work includes examining how the reverse VCs can be used most effectively.

Current work in the ATM Forum and ITU promises additional advantages for RSVP and ATM including renegotiating QoS parameters and variegated VCs. QoS renegotiation would be particularly beneficial since the only option available today for changing VC QoS parameters is replacing the VC. It is important to keep current with changes in ATM, and to keep this document up-to-date.

Scaling of the number of sessions is an issue. The key ATM related implication of a large number of sessions is the number of VCs and associated (buffer and queue) memory. The approach to solve this problem is aggregation either at the RSVP layer or at the ISSLL layer (or both).

This document describes approaches that can be used with ATM UNI4.0, but does not make use of the available leaf-initiated join, or LIJ, capability. The use of LIJ may be useful in addressing scaling issues. The coordination of RSVP with LIJ remains a research issue.

Lastly, it is likely that LANE 2.0 will provide some QoS support mechanisms, including proper QoS allocation for multicast traffic. It is important to track developments, and develop suitable RSVP over ATM LANE at the appropriate time.

8 Authors' Addresses


```
Steven Berson
USC Information Sciences Institute
4676 Admiralty Way
Marina del Rey, CA 90292

Phone: +1 310 822 1511
EMail: berson@isi.edu
```

Lou Berger
FORE Systems
6905 Rockledge Drive
Suite 800
Bethesda, MD 20817

Phone: +1 301 571 2534
EMail: lberger@fore.com

References

[1] Armitage, G., "Support for Multicast over UNI 3.0/3.1 based ATM Networks," Internet Draft, February 1996.

[2] The ATM Forum, "LAN Emulation Over ATM Specification", Version 1.0.

[3] The ATM Forum, "MPOA Baseline Version 1", 95-0824r9, September 1996.

[4] Berson, S., "'Classical' RSVP and IP over ATM," INET '96, July 1996.

[5] Borden, M., Crawley, E., Krawczyk, J, Baker, F., and Berson, S., "Issues for RSVP and Integrated Services over ATM," Internet Draft, February 1996.

[6] Borden, M., and Garrett, M., "Interoperation of Controlled-Load and Guaranteed-Service with ATM," Internet Draft, June 1996.

[7] Braden, R., Clark, D., Shenker, S. "Integrated Services in the Internet Architecture: an Overview," RFC 1633, June 1994.

[8] Braden, R., Zhang, L., Berson, S., Herzog, S., and Jamin, S., "Resource ReSerVation Protocol (RSVP) – Version 1 Functional Specification," Internet Draft, November 1996.

[9] Heinanen, J., "Multiprotocol Encapsulation over ATM Adaptation Layer 5," RFC 1483.

[10] Herzog, S., "Accounting and Access Control Policies for Resource Reservation Protocols," Internet Draft, June 1996.

[11] Laubach, M., "Classical IP and ARP over ATM," RFC 1577, January 1994.

[12] Luciani, J., Katz, D., Piscitello, D., Cole, B., "NBMA Next Hop Resolution Protocol (NHRP)," Internet Draft, June 1996.

[13] Onvural, R., Srinivasan, V., "A Framework for Supporting RSVP Flows Over ATM Networks," Internet Draft, March 1996.

[14] Perez, M., Liaw, F., Grossman, D., Mankin, A., Hoffman, E., and Malis, A., "ATM Signalling Support for IP over ATM," RFC 1755.

[15] Perez, M., Mankin, A. "ATM Signalling Support for IP over ATM - UNI 4.0 Update" Internet Draft, November 1996.

[16] *ATM User-Network Interface (UNI) Specification - Version 3.1*, Prentice Hall.

[17] Shenker, S., Partridge, C., Guerin, R., "Specification of Guaranteed Quality of Service," Internet Draft, August 1996.

[18] Wroclawski, J., "Specification of the Controlled-Load Network Element Service," Internet Draft, August, 1996.

[19] Zhang, L., Deering, S., Estrin, D., Shenker, S., Zappala, D., "RSVP: A New Resource ReSerVation Protocol," IEEE Network, September 1993.