Internet Draft
Expires September 25, 1997
File: draft-ietf-issll-atm-imp-guide-00.ps

L. Berger
FORE Systems
March 1996

RSVP over ATM Implementation Guidelines

March 25, 1997

Status of Memo

Abstract

This note presents specific implementation guidelines for running RSVP over ATM switched virtual circuits (SVCs). It presents requirements and specific guidelines for running over today's ATM networks. The general problem is discussed in [5]. Integrated Services to ATM service mappings are covered in [7].

Author's 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

This note discusses running IP over ATM in an environment where SVCs are used to support QoS flows and RSVP is used as the internet level QoS signaling protocol. The general issues related to running RSVP[8] over ATM have been covered in several papers including [5, 4, 6, 11]. This document is intended as a companion to [5] and as a guide to implementers. The reader should be familiar with[5]. This document will define specific baseline requirements for implementations using ATM UNI3.x and 4.0. Some stated requirements must be adhered to by all RSVP over ATM implementations. Other stated requirements provide a baseline set of functionality, while allowing for more sophisticated approaches. We expect some vendors to additionally provide some of the more sophisticated approaches described in [5], and some networks to only make use of such approaches. The baseline set of functionality is defined to ensure predictability and interoperability between different implementations. We expect that the baseline requirements may change in the future, and at such a time this document will be replaced.

The rest of this section will define terms and assumptions used in the document. Section 2 will cover implementation guidelines specific to multicast sessions. Section 3 will cover implementation guidelines common to all RSVP session. Section 5 will conclude with a summary of stated requirements.

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:

- **RSVP** We assume RSVP as the internet signalling protocol which is described in [8]. The reader is assumed to be familiar with [8].
- **IPv4 and IPv6** RSVP support has been defined for both IPv4 and IPv6. The guidelines in this document are intended to be used to support RSVP with either IPv4 or IPv6. This document does not require on version over the other.
- **Best effort service model** The current Internet only supports best effort service. We assume that as additional components of the Integrated Services model that best effort service will continue to be a supported.
- **ATM UNI 3.x and 4.0** We assume ATM service as defined by UNI 3.x and 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 support issues are similar to point-to-point links. This draft assumes that SVCs are used to support RSVP over ATM.

2 Multicast RSVP Session Support

There are several aspects to running RSVP over ATM that are particular to multicast sessions. These issues result from the nature of ATM point-to-multipoint connections. This section addresses multicast end-point identification, multicast data distribution, multicast receiver transitions and next-hops requesting different QoS values (heterogeneity) which includes the handling of multicast best-effort receivers. Handling of best-effort receivers is not strictly an RSVP issues, but needs to be addressed in any RSVP over ATM implementation in order to maintain expected Internet service. Implementation guidelines for issues related to all RSVP sessions are covered in Section 3. Some of these guidelines cover issues that have special interactions for multicast session, these interactions are covered together with the more general issues.

2.1 Data VC Management for Heterogeneous Sessions

The issues relating to data VC management of heterogeneous sessions are covered in detail in [5] and not repeated. In summary, heterogeneity occurs when receivers request different levels of QoS within a single session, and also when some receivers do not request any QoS. Both types of heterogeneity are shown in figure 1.

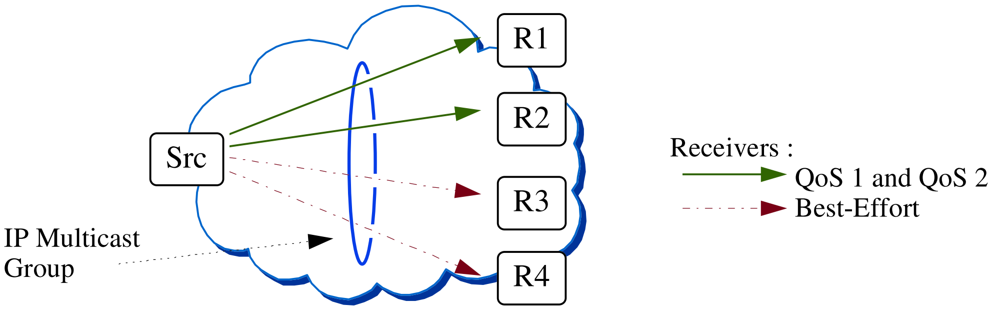

Figure 1: Types of Multicast Receivers

[5] provided four models for dealing with heterogeneity: full heterogeneity, limited heterogeneity, homogeneous, and modified homogeneous models. 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.

The key issue to be addressed by an implementation is providing requested QoS downstream. One of or some combination of the discussed models [5] may be used to provide requested QoS. Unfortunately, none of the described models is the right answer for all cases. For some networks, e.g. public WANs, it is likely that the limited heterogeneous model or a hybrid limited-full heterogeneous model will be desired. In other networks, e.g. LANs, it is likely that a the modified homogeneous model will be desired.

Since there is not one model that satisfies all cases, implementations must implement one of either the limited heterogeneity model or the modified homogeneous model. Implementations should support both approaches and provide the ability to select which method is actually used, but are not required to do so. Implementations may also support heterogeneity through some other mechanism, e.g., using multiple appropriately sized VCs.

2.2 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 usually provide QoS end-points, but not best-effort end-points.

There is even a case where RSVP next-hop information will not provide the appropriate end-point. This occurs when the next-hop is not RSVP capable, and RSVP is being automatically tunneled. 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 host case, MARS can 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. To ensure proper behavior, baseline RSVP over ATM implementations must only establish RSVP-initiated VCs to RSVP capable end-points. It is permissible to allow a user to override this behavior.

2.3 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 2 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.

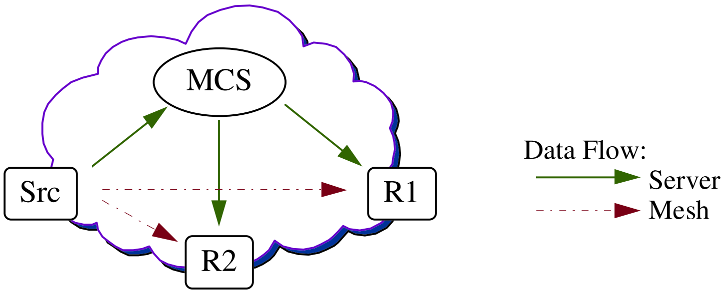

Figure 2: IP Multicast Data Distribution Over ATM

Current multicast servers [1] do not support any mechanisms for communicating QoS requirements to a multicast server. For this reason, RSVP over ATM implementations must support "mesh-mode" distribution for RSVP controlled multicast flows. When using multicast servers that do not support QoS requests, a sender must set the service, not global, break bit(s).

In the case of MARS[1], the selection of distribution modes is administratively controlled. Therefore network administrators that desire proper RSVP over ATM operation must appropriately configure their network to support mesh mode distribution for multicast groups that will be used in RSVP sessions.

2.4 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 must not be sent on a new VCs until 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 must be completed first, then the drop. This last behavior requires receivers to be prepared to receive some duplicate packets at times of QoS setup.

3 General RSVP Session Support

This section provides implementation guidelines that are common for all (both unicast and multicast) RSVP sessions. The section covers RSVP VC usage, QoS VC initiation, VC teardown, reservation to VC Mapping, handling requested changes in QoS, short-cuts, and encapsulation.

3.1 RSVP Message VC Usage

[5] covered the issues related to VC usage by RSVP messages. It discussed several options including: mixed control and data, single control VC per session, single control VC multiplexed among sessions, and multiple VCs multiplexed among sessions. QoS for control VCs was also discussed. Again that discussion is not repeated, [5] should be reviewed for detailed information.

Baseline RSVP over ATM implementations must send RSVP control (messages) over the best effort data path, see figure 3. It is permissible to allow a user to override this behavior. The stated approach minimizes VC requirements since the best effort data path will need to exist in order for RSVP sessions to be established and in order for RSVP reservations to be initiated. The specific best effort paths that will be used by RSVP are: for unicast, the same VC used to reach the unicast destination; and for multicast, the same VC that is used for best effort traffic destined to the IP multicast group. Note that for multicast there may be another best effort VC that is used to carry session data traffic, i.e., for data that is both in the multicast group and matching a sessions protocol and port.

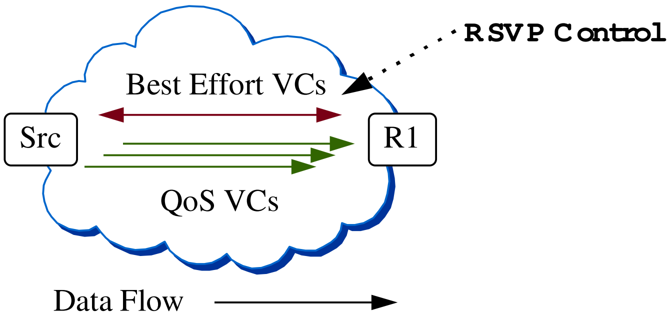

Figure 3: RSVP Control Message VC Usage

The disadvantage of this approach is that best effort VCs may not provide the reliability that RSVP needs. However the best-effort path is expected to satisfy RSVP reliability requirements in most networks. Especially since RSVP allows for a certain amount of packet loss without any loss of state synchronization.

3.2 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 must 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 4 shows this model of data flow VC initiation.

Baseline RSVP over ATM implementations are prohibited from sending data on a RSVP initiated QoS VC in the backwards direction. There are two reasons. The first is that use of the backwards data path requires the VC initiator to appropriate set backwards QoS parameters. The second is

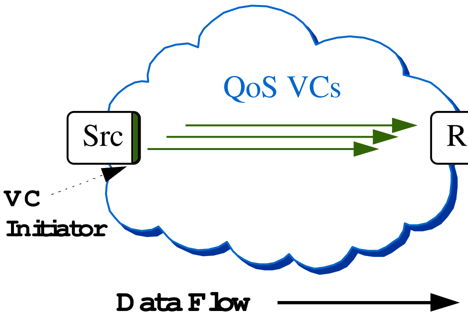

Figure 4: Data Flow VC Initiation

that backwards data paths are not available with point-to-multipoint VCs, so backwards data paths could only be used to support unicast RSVP reservations. Since reverse path usage is a special case that cannot be used to support multicast flows and such use for unicast requires some undefined out of band communication, implementations must not send data on QoS VCs in the backwards direction.

3.3 VC Teardown

Normally data VCs are torndown based on inactivity timers. This mechanism is used since IP is connectionless and there is therefore no way to know when a VC is no longer needed. Since RSVP provides explicit mechanisms (messages and timeouts) to determine when an associated data VC is no longer needed, the traditional VC timeout mechanisms is not needed. Data VCs set up to support RSVP controlled flows should only be released at the direction of RSVP. Such 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[12], 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[13] states that inactivity timers must not be used at the VC receiver.

In RSVP over ATM implementations, the configurable inactivity timer mentioned in [12] must be set to “infinite” for VCs initiated at the request of RSVP. Setting the inactivity timer value at the VC initiator should not be problematic since the proper value can be relayed internally at the originator. Setting the inactivity timer at the VC receiver is more difficult, and would require some mechanism to signal that an incoming VC was RSVP initiated. To avoid this complexity and to conform to [13], RSVP over ATM implementations must not use an inactivity timer to clear any received connection.

3.4 Reservation to VC Mapping

As discussed in [5], data associated with multiple RSVP sessions could be sent using the same shared VCs. Implementation of such "aggregation" models is still a matter for research. Therefore, baseline RSVP over ATM implementations must use independent VCs for each RSVP session. Implementations may also support aggregation approaches. Use of such approaches must be at the discretion of the user.

3.5 Dynamic QoS

As stated in [5], there is a mismatch in the service provided by RSVP and that provided by ATM UNI3.x and 4.0. RSVP allows modifications to QoS parameters at any time, while ATM does not support any modifications to QoS parameters after VC setup. See [5] for more detail.

The baseline 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 only then is the old VC 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. Implementations may also limit number of changes processed in a time period per [5].

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.

3.6 Short-Cuts

Short-cuts [10] 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 5, PATH and RESV messages following different paths.

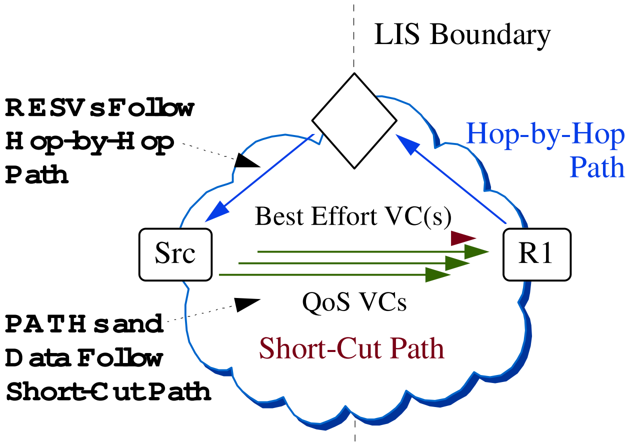

Figure 5: Asymmetric RSVP Message Forwarding With ATM Short-Cuts

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.

The key issue to be addressed by any RSVP over ATM solution is when to establish a short-cut for a QoS data flow. Baseline RSVP over ATM implementations should simply follow best-effort traffic. When a short-cut has been established for best-effort traffic to a destination or next-hop, that same end-point should be used when setting up RSVP triggered VCs for QoS traffic to the same destination or next-hop. This will happen naturally when PATH messages are forwarded over the best-effort short-cut. Note that in this approach when best-effort short-cuts are never established, RSVP triggered QoS short-cuts will also never be established.

3.7 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 multiple encapsulation options for running IP over ATM, for example RFC 1483[9] and LANE[2]. There is also other encapsulation options, such as MPOA[3].

Baseline RSVP over ATM implementations must use a consistent encapsulation scheme for all IP over ATM packets. This includes RSVP packets and associated IP data packets. So, encapsulation used on QoS data VCs and related control VCs must be the same as used by best-effort VCs.

4 Security

The same considerations stated in [8] and [12] apply to this document. There are no additional security issues raised in this document.

5 Implementation Summary

This section provides a summary of previously stated requirements.

5.1 Requirements

All RSVP over ATM UNI 3.0 and 4.0 implementations must conform to the following:

- **Heterogeneity**
  Implementations must not, in the normal case, send more than one copy of a particular data packet to a particular next-hop (ATM end-point).
  Implementations must ensure that data traffic is sent to best-effort receivers.

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

- **Receiver Transitions**
  While creating new VCs, senders must either send on only the old VC or on both the old and the new VCs.

- **VC Initiation**
  All RSVP triggered QoS VCs must be established by the sub-net senders.
  VC receivers must be able to accept incoming QoS VCs.

- **VC Teardown**
  VC initiators must not tear down RSVP initiated VCs due to inactivity.
  VC receivers must not tear down any incoming VCs due to inactivity.

5.2 Baseline Requirements

Baseline requirements define a minimum set of functionality that must be provided by implementations. Implementations may also provide additional functionality that may be configured to override the baseline behavior. Which behavior is selected is a policy issue for network providers. We expect some networks to only make use of baseline functionality and others to only make use of additional functionality.

- **Heterogeneity**
  Either limited heterogeneity model or the modified homogeneous model must be supported for handling heterogeneity.
  Implementations should support both approaches and provide the ability to select which method is actually used, but are not required to do so.

- **Multicast End-Point Identification**
  Implementations must only establish RSVP-initiated VCs to RSVP capable end-points.

- **Multicast Data Distribution**
  Implementations must support "mesh-mode" distribution for RSVP controlled multicast flows. In the case of MARS, distribution mode is administratively controlled. So this requirement can only be implemented by network administrators.

- **RSVP Control VC Management**
  Implementations must send RSVP control (messages) over the best effort data path.

- **Reservation to VC Mapping**
  Implementations must use a single VC to support each RSVP reservation.

- **Dynamic QoS**
  Implementations must support RSVP requested changes in reservations by attempting 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.
- **Short-Cuts**
  Implementations should establish QoS short-cut whenever a best-effort short-cut is in use to a particular destination or next-hop. This means that when best-effort short-cuts are never established, RSVP triggered short-cuts also should not be established.
- **Encapsulation**
  Implementations must encapsulate data sent on QoS VCs with the same encapsulation as is used on best-effort VCs.

6 Author's Address

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] Berson, S., Berger, L., "IP Integrated Services with RSVP over ATM," Internet Draft, March 1997.

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

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

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

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

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

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