Network Working Group
Request for Comments: 1458

R. Braudes
S. Zabele
TASC

May 1993

Requirements for Multicast Protocols

Status of this Memo

This memo provides information for the Internet community. It does not specify an Internet standard. Distribution of this memo is unlimited.

Summary

Multicast protocols have been developed over the past several years to address issues of group communication. Experience has demonstrated that current protocols do not address all of the requirements of multicast applications. This memo discusses some of these unresolved issues, and provides a high-level design for a new multicast transport protocol, group address and membership authority, and modifications to existing routing protocols.

Table of Contents

1. Introduction

Multicast protocols have been developed to support group communications. These protocols use a one-to-many paradigm for transmission, typically using class D Internet Protocol (IP) addresses to specify specific multicast groups. While designing network services for reliable transmission of very large imagery as part of the DARPA-sponsored ImNet program, we have reviewed existing multicast protocols and have determined that none meet all of the requirements of image communications [3]. This RFC reviews the current state of multicast protocols, highlights the missing features, and motivates the design and development of an enhanced multicast protocol.

First, the requirements for network services and underlying protocols related to image communications are presented. Existing protocols are then reviewed, and an analysis of each protocol against the requirements is presented. The analyses identify the need for a new multicast protocol. Finally, the features of an ideal reliable multicast protocol that adapts to network congestion in the transmission of large data volumes are presented. Additional network components needed to fully support the new protocol, including a Multicast Group Authority and modifications to existing routing protocols, are also introduced.

2. The Image Communications Problem

2.1 Scope

Image management and communications systems are evolving from film-based systems toward an all-digital environment where imagery is acquired, transmitted, analyzed, and stored using digital computer and communications technologies. The throughput required for communicating large numbers of very large images is extremely large, consisting of thousands of terabytes of imagery per day. Temporal requirements for capture and dissemination of single images are stringent, ranging from seconds to at most several minutes. Imagery will be viewed by hundreds of geographically distributed users who will require on-demand, interactive access to the data.

Traditional imaging applications involve images on the order of 512 by 512 pixels. In contrast, a single image used for remote sensing can have tens of thousands of pixels on a side. Multiplying the data volume associated with remotely sensed images by even a small number of users clearly motivates moving beyond the current suite of reliable protocols.

Basic image communication applications involve distribution of individual images to multiple users for both individual and collaborative analyses, and network efficiency requires the use of multicast protocols. Areas where multicasting offers significant advantages include real-time image acquisition and dissemination, distribution of annotated image-based reports, and image conferencing. Images are viewed on a heterogeneous set of workstations with differing processing and display capabilities, traveling over a heterogeneous network with bandwidths varying by up to six orders of magnitude between the initial down link and the slowest end user.

2.2 Requirements

Multicast protocols used for image communications must address several requirements. Setting up a multicast group first requires assigning a multicast group address. All multicast traffic is then delivered to this address, which implies that all members of the group must be listening for traffic with this address.

Within an image communications architecture such as that used for the ImNet program, diversity and adaptability can be accommodated by trading quality of service (i.e., image quality) with speed of transmission. Multicast support for quality-speed trades can be realized either through the use of different multicast groups, where each group receives a different image quality, or through the use of a single hierarchical stream with routers (or users) extracting relevant portions.

Due to the current inability of routers to support selective transmission of partial streams, a multiple stream approach is being used within ImNet. Efficient operation using a multiple stream approach requires that users be able to switch streams very quickly, and that streams with no listeners not be disseminated. Consequently, rapid configuration of multicast groups and rapid switching between multicast groups switching is essential.

Inevitably, network congestion or buffer overruns result in packet loss. A full range of transport reliability is required within an image communications framework. For some applications such as image conferencing, packet loss does not present a problem as dropped mouse

movements can be discarded with no meaningful degradation in utility. However, for functions such as image archiving or detailed image analysis, transport must be completely reliable, where any dropped packets must be retransmitted by the sender. Additionally, several hierarchical image compression methods can provide useful, albeit degraded, imagery using a semi-reliable service, where higher level data is transmitted reliably and the lower level data is transmitted unreliably.

In support of reliable transport, image communications services must also support adaptation to network congestion using flow control mechanisms. Flow control regulates the quantity of data placed on the network per unit time interval, thereby increasing network efficiency by reducing the number of dropped packets and avoiding the need for large numbers of retransmissions.

3. Review of Existing Multicast Protocols

Several existing protocols provide varying levels of support for multicasting, including IP/Multicast [5], the Xpress Transfer Protocol (XTP) [11], and Experimental Internet Stream Protocol Version 2 (ST-II) [10]. While the Versatile Message Transaction Protocol (VMTP) [4] also supports multicast, it has been designed to support the transfer of small packets, and so is not appropriate for large image communications. Additionally, a specification exists for the Multicast Transport Protocol (MTP) [2].

The image communication requirements for a multicast protocol include multicast group address assignment, group set-up, membership maintenance (i.e., join, drop, and switch membership), group tear-down, error recovery, and flow control, as presented above. The remainder of this section discusses how well each of the existing protocols meets these requirements.

3.1 IP/Multicast

IP/Multicast is an extension to the standard IP network-level protocol that supports multicast traffic. IP/Multicast has no address allocation mechanism, with addresses assigned either by an outside authority or by each application. This has the potential for address contention among multiple applications, which would result in the traffic from the different groups becoming commingled.

There is no true set-up processing for IP/Multicast; once an address is determined, the sender simply transmits packets to that address with routers determining the path(s) taken by the data. The receiver side is only slightly more complex, as an application must issue an add membership request for IP to listen to traffic destined to the

desired address. If this is the first member of a group, IP multicasts the request to routers on the local network using the Internet Group Multicast Protocol (IGMP) for inclusion in routing tables. Multicast packets are then routed like all other IP packets, with receivers accepting traffic addressed to joined groups in addition to the normal host address.

A major problem with the IP/Multicast set-up approach is informing hosts of multicast group addresses. If addresses are dynamically allocated, then a mechanism must be established for informing receivers which addresses have been assigned to which groups. This requires a minimum of one round trip time, with an address requested from a server and then returned to the receiver.

Dropping membership in a group involves issuing a request to the local IP, which decrements the count of members in the IP tables. However, no special action is taken when group membership goes to zero. Instead, a heartbeat mechanism is used in which hosts are periodically polled for active groups, and routers stop forwarding group traffic to a network only after several polls receive no activity requests for that group to ensure that a membership report is not lost or corrupted in transit. This causes the problem of unneeded traffic being transmitted, due to a long periodicity for the heartbeat (minimum of one minute between polls); consequently there is no method for quickly dropping a group over a given path, impeding attempts to react to network congestion in real-time.

Finally, there is no transport level protocol compatible with IP/Multicast that is both reliable and implements a flow control mechanism.

3.2 XTP

XTP is a combined network and transport level protocol that offers significant support for multicast transfers. As with IP/Multicast, XTP offers no inherent address management scheme, so that an outside authority is required.

XTP is also similar to IP/Multicast as there is no explicit set-up processing between the sender and the receivers prior to the establishment of group communications. While there is implicit processing in key management, an external mechanism is required for passing the multicast group address to the receivers. The receivers must have established "filters" for the address prior to transmission in order to receive the data, and suffers the same problems as IP/Multicast.

In contrast to IP/Multicast, XTP does require explicit handshaking

between the sender and receivers that wish to join an existing group; however, there is no parallel communication for receivers dropping out of groups, and the only mechanism for a sender to know if there are any receivers is the polling scheme used for error control and recovery. This causes the same problems with sending traffic to groups without members discussed under IP/Multicast.

The XTP specification does not address how routers distribute a multicast stream among different connected networks; however it does include a discussion of the optional bucket, damping, slotting, and cloning algorithms to reduce duplicate multicast traffic within a local network.

The specification allows the user to determine whether multicast transfers are unreliable or semi-reliable, where semi-reliable transfers are defined to provide a "high-probability of success [9]" of delivery to all receivers. Reliability cannot be guaranteed due to the fact that XTP does not maintain the cardinality of the receiver set, and so cannot know that the data has been received by all hosts.

XTP recovers from errors using a go-back-n approach (assuming that the bucket algorithm has been implemented) by retransmitting dropped packets to all members of the multicast group, as group members are unknown. This has the potential of flooding the network if only a single receiver dropped a packet. If all dropped packets belong to a single network on an internet, with traffic generated over the entire connected network.

3.3 ST-II

ST-II is another network protocol that provides support for multicast communications. Similar to IP/Multicast and XTP, ST-II requires a separate application-specific protocol for assigning and communicating multicast group addresses.

While ST-II is a network level protocol, it guarantees end-to-end bandwidth and delay, and so obviates the need for many of the functions of a transport protocol. The guarantee is provided by requiring bandwidth reservations for all connections, which are made at set-up time, and ensuring that the requested bandwidth is available throughout the lifetime of the connection. The enforcement policy ensures that the same path is followed for all transmissions, and prohibits new connections over the network unless there is sufficient bandwidth to accommodate the expected traffic. This is accomplished by maintaining the state of all connections in the network routers, trading the overhead of this connection set-up for the performance guarantees.

Connection set-up involves negotiation of the bandwidth and delay parameters and path between the sender, intermediate routers, and receivers. If the requested resources cannot be made available, the sender is given the option of either accepting what is available or canceling the connection request.

To add a new user to an existing group, the new receiver must first communicate directly with the sender using a different protocol to exchange relevant information such as the group address. The sender then requests ST-II to add the new receiver, with the basic connection set-up processing invoked as before with the new connection completed only if there is sufficient bandwidth to process the user.

While the resource guarantee system imposed by ST-II tries to prevent network congestion from occurring, there are situations where priority traffic must be introduced into the network. ST-II makes this very expensive, as the resource requirements for existing connections must be adjusted, which can only be accomplished by the origin of each stream. This must be completed prior to the connection set-up for the priority stream, introducing a large delay before the important data can be transmitted.

ST-II connections can be closed by either the sender or the receiver. When the last receiver along a path has been removed, the resources allocated over that path are released. When all receivers have been removed, the sender in informed and has the option of either adding a new receiver or tearing down the group.

3.4 MTP

MTP is a transport level protocol designed to support efficient, reliable multicast transmissions on top of existing network protocols such as IP/Multicast. It is based on the notion of a multicast "master" which controls all aspects of group communications.

Allocation of a specific group address, or network service access point, is not considered part of MTP, and as with the other multicast protocols requires the use of an outside addressing authority. The MTP specification does require the master to make a "robust effort [2]" to ensure the address selected is not already in use by trying to join an existing group at that address, but the problems described above remain.

Once the address is established, receivers issue a request to join the existing group using a unique connection identifier that is pre-assigned. The MTP specification addresses neither how the identifier is allocated nor how the receivers learn its value, but is assumed to

be handled through an external protocol. The join request specifies whether the receiver wishes to be a producer of information or only a receiver, whether the connection should be reliable or best effort, whether the receiver is able to accept multiple senders of information, the minimum throughput desired, and the maximum data packet size. If the request can be granted, then the master replies with an ACK with a multicast connection identifier; otherwise a NAK is returned.

Dropping membership in a group is coordinated through the master. The specification does not address what action the master should take when the group is reduced to a single member, but a logical action would be to stop distributing transmit tokens if there are no active receivers.

One of the major features in MTP is the ordering of received data. The master distributes transmit tokens to data producers in the group, which allow data to be provided at a specified rate. Rate control provides flow control within the protocol, with members that cannot maintain a minimum flow requested to leave the group.

Error recovery utilizes a NAK-based selective retransmission scheme. Senders are required to maintain data for a time period specified by the master, and to be able to retransmit this data when requested by members of the group. These retransmissions are multicast to the entire group, requiring receivers to be able to cope with duplicate packets. If a retransmission request arrives after the data has been released, the sender must NAK the request.

A potential problem with MTP is the significant amount of overhead associated with the protocol, with virtually all control traffic flowing through the master. The extra delay and congestion makes MTP inappropriate for the image dissemination applications.

3.5 Summary

Our analysis has determined that there are significant problems with all of the major multicast protocols for the reliable, adaptive multicast transport of large data items. The problems include inadequate address management, excessive processing of control information, poor response to network congestion, inability to handle high priority traffic, and suboptimal error recovery and retransmission procedures. We have developed a high-level notion of the requirements for a service that addresses these issues, which we now discuss.

4. Protocol Suite for Reliable, Adaptive Multicast

We present an integrated set of three basic components required to provide a reliable multicast service: the Multicast Group Authority (MGA); the Reliable, Adaptive Multicast Protocol (RAMP); and modified routing algorithms. These components are designed to be compatible with, and take full advantage of, reservation systems such as RSVP [12].

In this discussion, we have broadened the definition of the term "Quality of Service (QOS)." There are many applications where the information content of the underlying data can be reduced through data compression techniques. For example, a 1,024 x 1,024 pixel image can be sub-sampled down to 512 x 512 pixels. This degradation results in a lower quality of service for the end user, while reducing the traditional network QOS requirements for the transfer.

4.1 The Multicast Group Authority

The Multicast Group Authority (MGA) provides services related to managing the multicast address space and high-level management support to existing multicast groups. The MGA has three primary responsibilities: address management, service registration, and group membership maintenance.

The MGA is hierarchical in nature, similar to the Internet Domain Name System (DNS) [7]. Requests for service are directed to an MGA agent on the local workstation, which are propagated upwards as required.

4.1.1 Address Management

The MGA is responsible for the allocation and deallocation of addresses within the Internet Class D address space. Address requests received from application processes or other MGA nodes result in a block of addresses being assigned to the requesting MGA node. The size of the address block allocated is dependent on the position of the requester in the MGA hierarchy, to reduce the number of address requests propagated through the MGA tree.

Figure 1 can be used to show what happens when an application requests a multicast address from the authority at node 1.1.1. Assuming that this is the first request from this branch of the MGA, node 1.1.1 issues a request to its parent, node 1.1, which propagates the request to node 1. Node 1 passes this request to the root, which issues a block of, say, 30 class D addresses. Of these 30, 10 are returned to node 1.1, with the remaining 20 reserved for requests from node 1’s other children. Similarly, node 1.1 passes 3 addresses

to node 1.1.1, reserving the other 7 for future requests. Finally, node 1.1.1 answers the applications request for an address, keeping the remaining 2 addresses for future use.

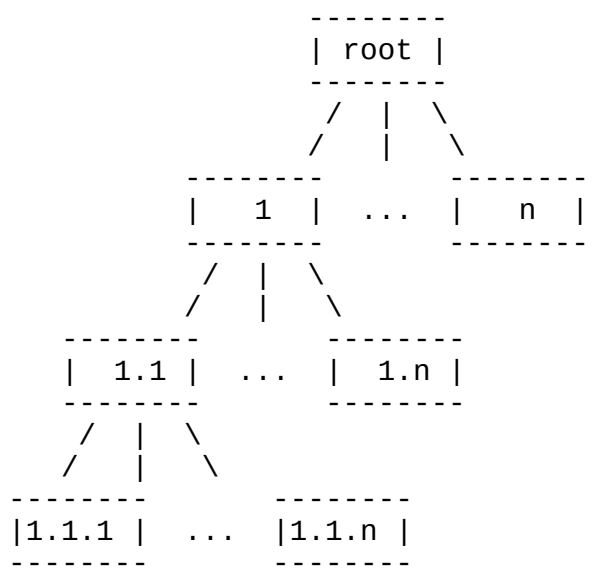

Figure 1. Sample MGA Hierarchy

When the root exhausts the address space, a request is made to the children for reclamation of unused addresses. This request propagates down the tree, with unused addresses passed back through the hierarchy and returned to the address pool. If the entire address space is in use, then requests for additional addresses are not honored.

When an application no longer requires an address, it is returned to the local MGA node, which keeps it until either it is requested by another application, it is requested by its parent, or the node is terminated. At node termination, all available addresses are returned to the parent. Parents periodically send heartbeat requests to their children to ensure connectivity, and local nodes similarly poll applications, with addresses recalled if the queries are not answered.

4.1.2 Service Registration, Requests, Release, and Group Membership Maintenance

The MGA maintains the state of all registered multicast services and receivers. State information includes the number of members associated with each group by requested QOS reliability, which is updated as services are offered or rescinded and as members join or

leave a group. The state information is used to ensure that there is at least one group member listening to each multicast transfer.

Servers register the availability of service, specifying whether reliable service is available [section 4.2.2] and optionally the number of qualities of service offered [section 4.2.1]. A multicast group address is allocated from the address pool and the service is assigned an identifier as required. If a reservation protocol that requires information from the server (such as RSVP) is in use, then the MGA notifies the reservation system of the service with any required parameters. The service registration is propagated through the MGA, so that potential clients can discover service availability. However, servers do not begin data transfers until directed to do so by the MGA.

Client requests for service are also processed through the MGA. Service requests specify a service, a desired quality of service, and a reliability indication. If the request is for a service that has been registered, then the routing support is directed to add a route for the new user [section 4.3.1]. If necessary, the MGA also notifies the reservation protocol. If either the requested QOS is not being provided or it is provided unreliably and the request is for reliable transport, then the service provider is also notified. If the service has not yet been registered, an identifier for the service is assigned and the request is queued for when the service is registered. In either case, a response is sent to the requester.

Requests for termination of group membership are also sent to the MGA. If the request originates at a client, the MGA notifies the routing function and reservation protocol of the termination in case the route should be released [section 4.3.2]. If termination results in a given QOS no longer having any recipients, the service provider is notified that the QOS is no longer required and should not be transmitted. Server-directed group terminations follow a similar procedure, with all clients of the group notified, and the service offering is removed from the MGA state tables.

4.2 The Reliable Adaptive Multicast Protocol (RAMP)

RAMP is a transport-level protocol designed to provide reliable multicast service on top of a network protocol such as IP/Multicast, with unreliable transport also available. RAMP is build on the premise that applications can request one quality of service (using our extended definition), but only require reliable transmission at a lower level of quality. For example, consider the transmission of hierarchical image data, in which a base spatial resolution is transmitted, followed by higher resolution data. An application may require the base data to be sent reliably, but can tolerate dropped

packets for the higher resolution by using interpolation or pixel replication from the base level to approximate the missing data. Similar methods can be applied to other data types, such as audio or video.

4.2.1 Quality of Service Levels

RAMP allows a multicast service to be provided at multiple qualities of service, with all or some of these levels transmitted reliably. These QOS can be distributed across different groups using different class D addresses, or in the simplest case be transmitted in individual groups. Single packets can be used for either a single QOS, or may be applicable to multiple qualities of service.

When a data packet is transmitted, a header field indicates the QOS level(s) associated with that packet. In the old IP implementations, the Type of Service field can be used as a bit field with one bit for each applicable QOS, although this is incompatible with RFC 1349 [1]. If a packet is required for multiple QOS, then multiple values are encoded in the field. The RAMP host receiver protocol only accepts those packets addressed to a group in which an application has requested membership and that has a QOS value which is in the set of values requested by the receivers.

The quality of service requested within a flow can be modified during the life of the flow. QOS modification requests are forwarded to the MGA, which reduces the number of receivers in the original QOS group and increments the count for the requested QOS. These changes are propagated through the MGA hierarchy, with the server notified if either the original QOS has no remaining receivers or if the new QOS is not currently being served; similarly, the routers are notified if routing changes are required.

4.2.2 Error Recovery

Sequence numbers are used in RAMP to determine the ordering of packets within a multicast group. Mechanisms for ordering packets transmitted from different senders is a current research topic [2, 6], and an appropriate sequencing algorithm will be incorporated within the protocol.

Applications exist that do not require in-order delivery of data; for example, some image servers include position identification information in each packet. To enhance the efficiency of such schemes, RAMP includes an option to allow out-or-order delivery of packets to a receiver.

A NAK-based selective retransmission scheme is used in RAMP to minimize the protocol overhead associated with ACK-based schemes. When a receiver notices that one or more packets have not been received, and the transmission is reliable, a request is sent to the sender for the span of packets which are missing.

RAMP at the sender aggregates retransmission requests for the time specified by the retransmission hold timer [section 4.2.3]. After this time, the requests are evaluated to determine if sufficient receivers dropped a given packet to make multicasting the retransmission worthwhile by comparing it to a threshold value. All packets that have received a number of retransmission requests greater than the threshold are multicast to the group address, with other packets unicast to the individual requesters. The proposed retransmission scheme is a compromise between the extremes of multicasting and unicasting all retransmissions. The rationale is that multicasting a request issued by a single sender unnecessarily floods networks which had no packet loss, while unicasting to a large number of receivers floods the entire network. The optimal approach, dynamically constructing a new multicast group for each dropped packet, is currently too costly in terms of group set-up time.

For those cases where the service provider is unable to retransmit the data due to released or overwritten buffers, the protocol delivers NAK responses using either multicast or unicast based on the number of retransmission requests received.

4.2.3 Flow Control

RAMP utilizes a rate-based flow control mechanism that derives rate reductions from requests for retransmission or router back-off requests (i.e., ICMP source quench messages), and derives rate increases from the number of packets transmitted without retransmission requests. When a retransmission request is received, the protocol uses the number of packets requested to compute a rate reduction factor. Similarly, a different reduction factor is computed upon receipt of a router-generated squelch request. The rate reduction factors are then used to compute a reduced rate of transmission.

When a given number of packets have been transmitted without packet loss, the rate of transmission is incrementally increased. The size of the increase will always be smaller than the size of the smallest rate decrease, in order to minimize throttling.

The retransmission hold timer is modified according to both application requests and network state. As the number of retransmission requests rises, the hold timer is incremented to

minimize the number of duplicate retransmissions. Similarly, the timer is decremented as the number of retransmission requests drops.

RAMP allows for priority traffic, which is marked in the packet header. The protocol transmits a variable number of packets from each sending process in proportion to the priority of the flow.

4.3 Routing Support

The protocol suite requires routing support for four functions: path set-up, path tear-down, forwarding based on QOS values, and prioritized packet loss due to congestion. The support must be integrated into routers and network-level protocols in a similar fashion to IGMP [8].

Partial support comes as a direct consequence of using reservation protocols such as RSVP. This RFC does not mandate the means of implementing the required functions, and the specified protocols are compatible with known reservation protocols.

The routers state tables must maintain both the multicast group address and the QOS level(s) requested for each group on each outbound interface in order to make appropriate routing decisions [section 4.3.3]. Therefore, the router state tables are updated whenever group membership changes, including QOS changes.

4.3.1 Path Set-up

Routers receive path set-up requests from the MGA as required when new members join a multicast group, which specifies the incoming and outgoing interfaces, the group address, and the QOS associated with the request. When the message is received, the router establishes a path between the server and the receiver, and subsequently updates the multicast group state table. The mechanism used to discern the network interfaces is not specified, but may take advantage of other protocols such as the RSVP path and reservation mechanism.

4.3.2 Path Tear-down

Path tear-down requests are also propagated through the routers by the MGA when group membership changes or QOS changes no longer require data to be sent over a given route. These are used to inform routers of both deletions of QOS for a given path and deletions of entire paths. The purpose of the message is to explicitly remove route table entries in order to minimize the time required to stop forwarding multicast data across networks once the path is no longer required.

4.3.3 Multicast Routing Based on Quality of Service

Traditional multicast routing formulates route/don't route decisions based on the destination address in the packet header, with packets duplicated as necessary to reach all destinations. In the proposed new protocol suite, routers also consult the QOS field for each packet as different paths may have requested different qualities of service. Packets are only forwarded if the group address has been requested and the quality of service specified in the header is requested in the state table entry for a given interface.

4.3.4 Quality of Service Based Packet Loss

Network congestion causes router queues to overflow, and as a result packet loss occurs. The QOS and priority indications in the packet headers can be used to prioritize the order in which packets are dropped. First, packets with the priority field set in the header are dropped last. Within packets of equal priority, packets are dropped in order of QOS, with the highest QOS packets dropped first. The rationale is that other packets with lower QOS may be usable by receivers, while packets with high QOS may not be usable without the lower QOS data.

5. Interactions Among the Components: An Example

The MGA, RAMP, and routing support functions all cooperate in the multicast process. As an example, assume that a network exists with a single server (S), three routers (R1, R2, and R3), and two clients (C1 and C2). The path between S and C1 goes through R1 and R2, while the path between S and C2 goes through R1, R2, and R3. The network is shown in figure 2.

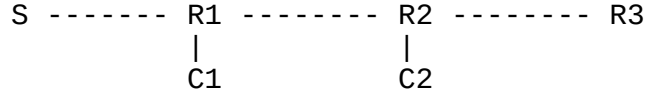

Figure 2. Sample Network Configuration

Service Registration

When S is initiated, it registers a service with the MGA node in the local workstation, offering reliable service at two qualities of service, Q1 and Q2. As this is the first multicast offering on the workstation, the local MGA requests a block of multicast addresses from the hierarchy, and assigns an address and service identifier to S. If the RSVP reservation protocol is in operation, the local MGA node in S notifies RSVP to send a RpathS message out for the service, which goes through R1, R2, and R3,

reaching the RSVP nodes on C1 and C2. The service and its characteristics are propagated throughout the MGA hierarchy, ultimately reaching the MGA nodes resident on C1 and C2. The service is now available throughout the network.

Service Request and Path Set-up

The client C1 requests reliable service from S at QOS Q1, by issuing a request to the MGA node in C1. If a reservation protocol is in use, then it is used to reserve bandwidth and establish a path between the sender and receiver, going through R1 and R2; otherwise, the path is established through R1 and R2 by the routing protocol. R1 now forwards all packets from S with QOS Q1 along the path to R2, which routes them to C1. In concert with the path set-up, the add membership request is propagated through MGA to the server workstation. The local MGA tables are checked and it is noted that the service is not currently being offered, so the server is notified to begin reliable distribution of the service at Q1.

Initial Delivery

The server now begins transmitting Q1 data which is observed by R1. R1 inspects the header and notes that the packet has QOS Q1. The routing tables specify that QOS Q1 for this address are only forwarded along the interface leading to R2, and R1 acts accordingly. Similarly, R2 routes the packet to C1. When the data arrives at C1, the RAMP node inspects the QOS and destination address fields in the header, accepts the packet, and forwards it to the C1 client process.

Error Recovery

During transmission, if the RAMP node in C1 realizes that packets have been dropped, a retransmission request is returned to the server identifying spans of the missing packets. The RAMP node accepts the packet, builds the retransmission packets, and sets the retransmission hold timer. When the timer expires, the number of retransmission requests for each missing packet is compared against the threshold, and the packets are either unicast directly to the requesters or multicast to the entire group. As in this case there is only requester, the threshold is not exceeded and the packets are retransmitted to C1Us unicast address.

Group Membership Addition

Client C2 now joins the group, requesting reliable transmission at QOS Q2. Following the process used for C1, the request propagates

through the MGA (and potentially reservation protocol) hierarchy. Upon completion of the request processing, R1 routes packets for QOS Q1 and Q2 to R2, while R2 forwards QOS Q1 packets to C1 and Q2 packets to R3; client C1 only accepts packets marked as Q1 while C2 only accepts Q2 packets. The server is notified that it now has clients for Q2, and begins serving that QOS in addition to Q1.

QOS Based Routing

First, assume that QOS Q1 data is independent of QOS Q2 data. When the server sends a packet with Q1 marked in the header, the packet is received by R1 and is forwarded to R2. R2 receives the packet, and sends it out the interface to C1, but not to R3. Next, the server delivers a packet for Q2. R1 receives the packet and sends it to R2, which forwards it to R3 but not to C1. R3 accepts the packet, and forwards it to C2.

Now, assume that either Q2 is a subset of Q1, or that receivers of Q1 data also require Q2 data as in conditional compression schemes. Therefore, all Q2 packets are marked for both Q1 and Q2, while the remaining Q1 packets only have Q1 set in the header. Q1-only packets are routed as before, following the path S -> R1 -> R2 -> C1. However, Q2 packets are now routed from S to R1 to R2, at which point R2 duplicates the packets and sends them to both C1 and R3, with R3 forwarding them to C2. At C1, these packets have Q1 marked, and so are accepted, while at C2 the packet is accepted as the Q2 bit is verified.

Group Membership Deletion

When C1 issues a drop membership request, the MGA on the client workstation is notified, and the request is propagated through the MGA hierarchy back to the server MGA node. In parallel, the routers are notified to close the path, as it is no longer required, possibly through the reservation protocol. As this is the last client for the Q1 QOS, the server is informed to stop transmitting Q1 data, with Q2 data unaffected. A similar process occurs when C2 drops membership from the group, leaving the server idle. At this point, the server has the option of shutting down and returning the group address to the MGA, or to continue in an idle state until another client requests service.

Acknowledgements

This research was supported in part by the Defense Research Projects Agency (DARPA) under contract number F19618-91-C-0086.

References

[1] Almquist, P., "Type of Service in the Internet Protocol Suite", RFC 1349, Consultant, July 1992.

[2] Armstrong, S., Freier, A., and K. Marzullo, "Multicast Transport Protocol", RFC 1301, Xerox, Apple, Cornell University, February 1992.

[3] Braudes, R., and S. Zabele, "A Reliable, Adaptive Multicast Service for High-Bandwidth Image Dissemination", submitted to ACM SIGCOMM '93.

[4] Cheriton, D., "VMTP: Versatile Message Transaction Protocol", RFC 1045, Stanford University, February 1988.

[5] Deering, S., "Host Extensions for IP Multicasting", STD 5, RFC 1112, Stanford University, August 1989.

[6] Mayer, E., "An Evaluation Framework for Multicast Ordering Protocols", Proceedings ACM SIGCOMM '92, Baltimore, Maryland, pp. 177-187.

[7] Mockapetris, P., "Domain Names - Concepts and Facilities," STD 13, RFC 1034, USC/Information Sciences Institute, November 1987.

[8] Postel, J., "Internet Control Message Protocol - DARPA Internet Program Protocol Specification", STD 5, RFC 792, USC/Information Sciences Institute, September 1981.

[9] Strayer, W., Dempsey, B., and A. Weaver, "XTP: The Xpress Transfer Protocol", Addison-Wesley Publishing Co., Reading, MA, 1992.

[10] Topolcic, C., Editor, "Experimental Internet Stream Protocol, Version 2 (ST- II)", RFC 1190, CIP Working Group, October 1990.

[11] "XTP Protocol Definition Revision 3.6", Protocol Engines Incorporated, PEI 92-10, Mountain View, CA, 11 January 1992.

[12] Zhang, L., Deering, S., Estrin, D., Shenker, S., and D. Zappala, "RSVP: A New Resource ReSerVation Protocol", Work in Progress, March 1993.

Security Considerations

Security issues are not discussed in this memo.

Authors' Addresses

Bob Braudes
TASC
55 Walkers Brook Drive
Reading, MA 01867

Phone: (617) 942-2000
EMail: rebraudes@tasc.com

Steve Zabele
TASC
55 Walkers Brook Drive
Reading, MA 01867

Phone: (617) 942-2000
EMail: gszabele@tasc.com