An Experimental Investigation of the Internet Integrated Services Model using the Resource Reservation Protocol (RSVP)

Mark Blackford and Dr. Jeffrey Austen
Department of Electrical and Computer Engineering
Tennessee Technological University
Cookeville, TN 38505
mjb8355@tntech.edu and jra1854@tntech.edu

David T. Marlow
Systems Research and Technology Department
Combat Systems Branch
Naval Surface Warfare Center, Dahlgren Division
Dahlgren, VA 22448
dmarlow@nswc.navy.mil

Abstract - **The Internet Protocol (IP) in use today was designed to offer network applications a best-effort delivery service for network traffic and no guarantees are made as to when or if the packets will be delivered to their final destination. Many new network applications need the network to meet certain delay and packet loss requirements, but the current best-effort delivery service does not have the ability to provide this level of service. The Internet Engineering Task Force has proposed the Internet Integrated Services (IIS) model to allocate network resources achieving the desired service and the resource ReSerVation Protocol (RSVP) to deliver requests on behalf of the application for these resources.**

In this paper the basic concepts of the IIS model and RSVP are described. An experiment is presented in which the network performance is evaluated, both with and without network resource allocation, under varying degrees of load with the purpose of assessing the benefits of resource allocation for particular data flows.

INTRODUCTION

The Internet Protocols [1] that are used today were designed to offer network applications a best-effort delivery service for network traffic. Best-effort service attempts to deliver data for applications, but there are no guarantees as to when, or if, the packets will be delivered to their final destination. This design works well for applications which are insensitive to network delays, such as remote login and file transfer programs, where time permits higher level protocols to guarantee reliable delivery of data. However, new network applications have been developed that require certain loss and delay bounds on packet delivery at the network layer. Multimedia, video and audio conferencing, virtual-reality and real-time applications perform well when tested on high speed local area networks. Yet, when these applications are used, possibly over great distances through many intermediate networks as typified by connection via the global Internet, the constant data flows created by these application often experience congestion which may cause unacceptably long delays and high packet loss. Because of the rapid growth of the Internet and the increased use of these new applications, network traffic is sure to increase and network performance is likely to degrade.

In this paper a review of the Internet Integrated Services (IIS) model is given along with a review of the delivery protocol for network resource reservation, RSVP. An introduction to the guaranteed Quality of Service (QoS) is also given followed by an experiment designed to assess the benefits of using the IIS model.

INTERNET INTEGRATED SERVICES MODEL

The proposed IIS model [2,3] offers a method of coordinating the use of network resources to achieve a QoS better than best-effort service. A data flow of packets for which best-effort service is insufficient can be transferred with a certain delay bound and delivery probability, known as a QoS level, while non-delay sensitive traffic can continue to use the best-effort service. The components of the IIS model include admission control, resource reservation, packet classification and scheduling, and policing.

Any application wishing to obtain a QoS level will have to go through additional steps beyond what is required for best-effort service. First, an application will make a request to the network admission control agent. Since network resources are limited, admission control is used to notify the application if the request for network resources can be supported. In addition, admission control can be used to give internet service providers the ability to turn down unauthorized requests for their network resources. Resource reservation protocols deliver requests for network resources to routers in the network. At each router, a packet classifier uses reservation and path state information to determine the QoS that each packet should receive; a packet scheduler uses this to control transmission priorities and, if necessary, determine which packets may be discarded. Policing is done to prevent data flows from using more resources than are reserved.

0-7803-3844-8/97/$10.00 © 1997 IEEE

THE RESOURCE RESERVATION PROTOCOL, RSVP

RSVP [4,5] fills the role of a resource reservation protocol. RSVP is a receiver oriented control protocol which makes reservations for unidirectional data flows. A sender application sends a path state message describing the traffic specifications (tspec) of its data flow. The path is stored by the routers as the message travels from the sender to the receiver. Upon receiving this message, a receiving application wishing to make a reservation for this data flow will transmit a reservation specification (rspec) message back along the same path to the sender indicating the type of service it desires. Routers along the path back to the sender apply admission control to the request, and if the request is granted the reservation message is forwarded to the next upstream router. Otherwise, a reservation error message is returned to the receiver.

RSVP uses "soft state" in the network routers to maintain reservations. The path state and the reservation state must be periodically refreshed with the corresponding messages from sending and receiving applications. This allows for graceful recovery from route changes and error conditions in addition to supporting dynamic membership changes in multicast groups. When a reservation is no longer needed, a terminate signal is sent from either the sender or the receiver notifying the routers to delete the state information for the flow. If the state is not refreshed within a set time limit, the state will time out and the reservation and path state is dropped.

Different reservation styles provide support for multicasting with RSVP. Reservation styles allow receiver applications to choose which data flows can use its reserved resources. A wild-card filter reservation allows any sender from a multicast group to use the reserved resources. Reservation messages for this type of filter must be forwarded back to every possible sender in a multicast group by the routers. A fixed filter reservation will allow only one sender to use the reserved resources. These reservation messages will only follow the return path to the sender specified by the filter. A receiver can also use a shared-explicit filter to allow multiple senders to use the same reservation. Reservation messages will only be forwarded back to the senders specified in the reservation.

THE GUARANTEED QOS

The proposed guaranteed QoS [6] offers network applications the ability to set a limit on the delay experienced by the packets in a received data flow. This service is initiated by a source application notifying the network components along the data path and the receiving application(s) of the flow's traffic characteristics. This is accomplished by sending a RSVP path state message containing the tspec of the data flow to the routers and receiver(s).

Since the allocation of resources for a data flow is affected by the burstiness of the flow, a mechanism is needed to control the rate at which traffic is injected into a network. This can be accomplished using the token bucket concept [7]. For each data flow there exists an imaginary bucket which can hold tokens for sending data. Tokens accumulate in the bucket at a specified token bucket rate. For a host to transmit a packet, the quantity of tokens in the bucket must be equal to or greater than the size of the packet. When a packet is transmitted, the corresponding number of tokens is removed from the bucket. If enough tokens do not exist in the bucket, a packet cannot be sent and the host must wait until there are enough tokens to send the packet. To prevent "hoarding", the bucket depth limits the maximum amount of tokens that a particular data flow may accumulate. Any further tokens that are due a flow with a full bucket are discarded. Burst of traffic are controlled by a peak rate which is the maximum rate at which the bucket may be emptied.

The tspec from a sender application consists of the token bucket parameters, a minimum policed unit size, and a maximum packet size. The token bucket has a bucket depth, b, measured in bytes, a bucket rate, r, measured in bytes per second and a peak rate, p, also measured in bytes per second. All bytes in a packet are counted, including the IP header, when determining peak and bucket rates. Over any time interval, T, the amount of data sent cannot exceed $M+pT$ where M is the maximum packet size in bytes. Finally, m is the minimum policed unit measured in bytes; any packets smaller that m will be policed as size m.

The reservation specification for the guaranteed QoS consists of a rate R, measured in bytes per second. R represents the amount of bandwidth that a receiver desires for the data flow.

Assuming that the data flow serviced by the guaranteed QoS is conformant to its traffic specification, the end-to-end delay is upper bounded by [6]

$$\left[\frac{b-m}{R}\cdot\frac{p-R}{p-r}\right]+\frac{M+C_{tot}}{R}+D_{tot}$$

for $r \le R < p$ and

$$\frac{M+C_{tot}}{R}+D_{tot}$$

for $r \le p \le R$ where C_{tot} is the rate dependent network delay measured in bytes and D_{tot} is rate independent delay measured in units of microseconds. The parameters C_{tot} and

D_{tot} are determined by the network and made available to the applications making reservations.

EXPERIMENTAL INVESTIGATION

The resource allocation capabilities described should benefit applications needing them when the interconnection of participants is over networks experiencing congestion. In order to assess the benefits of resource allocation, an experiment is being conducted through a joint effort by Tennessee Technological University and the Naval Surface Warfare Center, Dahlgren Division.

To investigate the IIS model, a network configuration must be used which will exploit the benefits of the using the model. The strategy proposed here is to use a network configuration that has a natural congestion point. Since most of the model is implemented within routers, the minimal configuration must have at least one router. However, to cause an appropriate bottleneck, a two router configuration is used as shown in Figure 1.

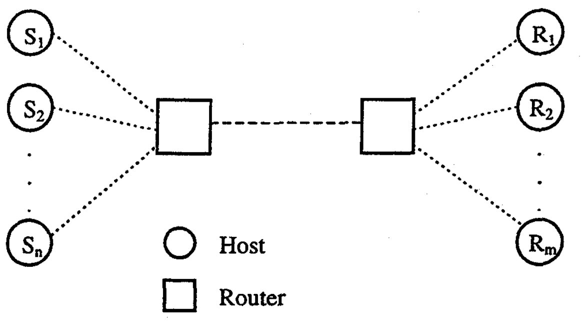

Figure 1. Proposed Test Configuration.

The hosts labeled S_1 through S_n are senders and hosts labeled R_1 through R_m are the receivers. The hosts are high performance workstations which are connected to the routers through a 155 Mb/s Asynchronous Transfer Mode (ATM) link. Workstation to workstation throughput over the ATM link has been measured to be approximately 56 Mb/s. The routers are using beta versions of the RSVP code and are connected to each other via a 10 Mb/s Ethernet link. Therefore, congestion can easily be caused in the link between the routers. Rather than use separate host computers for senders, multiple processes on one host perform the same function. Similarly, multiple processes on another host are used for receivers.

To determine baseline performance, measurements of packet delay and loss will be taken without any other network load for two cases: a best-effort flow and an RSVP flow. Then, with increasing degrees of network load the measurements will be repeated for both cases resulting in a set of data describing the characteristics of each flow in the presence of congestion.

It is expected the RSVP flow will experience similar delays and packet loss as compared to the best-effort service when the network is not experiencing congestion. As congestion increases, the RSVP case is expected to reach the level of delay that is guaranteed and become stable, while the best-effort flow should experience increasing delays and packet losses. The resulting data should give a good indication of the benefits of the IIS model.

SUMMARY

The Internet Integrated Services model provides a method for reserving network resources in packet switched networks. Using these reservations, the network can make guarantees on the delay that a flow will experience. In the future, the worldwide Internet will implement the IIS model and RSVP which will provide the various levels of network performance needed by an assortment of current and new real-time applications. Using the methods presented, the benefits of using Integrated Services over the current best-effort service can be assessed.

REFERENCES

[1] D. Comer and S. Stevens, Internetworking With TCP/IP Volume I, 2nd Ed., Englewood Cliffs, NJ, Prentice Hall, 1991.

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

[3] S. Shenkar and J. Wroclawski, "Network Element Service Specification Template," IETF Internet Draft, Version 4, Work in progress, November 1996.

[4] R. Braden et al., "Resource Reservation Protocol (RSVP) - Version 1 Functional Specification," IETF Internet Draft, Version 14, Work in progress, 5 November 1996.

[5] C. Huitema, Routing in the Internet, Englewood Cliffs, NJ, Prentice Hall, 1995.

[6] S. Shenkar and C. Partridge, "Specification of Guaranteed Quality of Service," IETF Internet Draft, Version 6, Work in progress, 13 August 1996.

[7] C. Partridge, "A Proposed Flow Specification," RFC 1363, September 1992.