An Efficient Adaptation of RSVP-TE in GMPLS

Zhiyu Zhou
Bell Labs Research China
Beijing 100080, China
zhiyuzhou@lucent.com

Donghui Gao
Electronic Engineering Department
Tsinghua University
gaodh02@mails.tsinghua.edu.cn

Keywords: MPLS, GMPLS, RSVP-TE, LDP, CR-LDP, LSP, LSR, OXC

Abstract

RSVP is extended to RSVP-TE to support LSP establishment in MPLS. Since the original RSVP was designed as the signaling protocol for IntServ QoS model, its receiver-initiated resource reservation shows a poor performance in MPLS, especially in the lightpath setup in optical networks via GMPLS. Based on the optical networks, this paper proposes a new adaptation of RSVP-TE. Both performance analyses and experimental results show this new scheme is superior to the original approach when the network size increases.

INTRODUCTION

In the early stage of the Internet, the IP network protocol offers the point-to-point best-effort service model for the simple and non time-critical applications such as e-mail, remote login, and ftp. With the commercial development of the Internet in the mid-1990s, many new applications such as remote video and multimedia conferencing asked for services of different qualities, which drove the Resource ReSerVation Protocol (RSVP) proposed for IntServ QoS model [1]. By 1997, RSVP has become a proposed standard and been widely implemented in IP networking equipment. However, RSVP has not been widely used in service provider networks because operators concern about its scalability and the overhead required to support potentially millions of host-to-host flows [2].

At the same time, Multi-Protocol Label Switching (MPLS) emerged from the IETF's effort to standardize a number of proprietary multilayer switching solutions and offer simpler mechanisms for packet-oriented traffic engineering and multiservice functionality with the added benefit of greater scalability [3]. MPLS separates the control plane from the forwarding plane logically, and encompasses signaling, path selection, and routing components in the control plane to support traffic engineering. Therefore, it is natural for the initial implementers of MPLS to extend RSVP for the signaling protocol of the control plane. RSVP provides both resource reservation that is an important component of traffic engineering and extensible mechanism of object carrying that can be easily used for explicit routing and label distribution of MPLS-TE.

In 1999-2000, the fast development of optical networks attracted many start-up companies and Internet Service Providers (ISPs) to invest in optical control plane signaling to stand ahead in the competition. The diversity and complexity in managing optical devices have driven the evolution and enhancement of MPLS into Generalized MPLS (GMPLS) in the IETF for label switching in not only packet-based domains, but also time, wavelength, and space domains. The evolution of MPLS into GMPLS has extended the signaling protocols including RSVP with Traffic Engineering extension (RSVP-TE) and Constraint-Based Routing Label Distribution Protocol (CR-LDP), and routing protocols such as Traffic Engineering Enhancements to Open Shortest Path First (OSPF-TE) and Intermediate System to Intermediate System (IS-IS-TE). Moreover, Automatically Switched Optical Networks (ASON) of ITU-T has employed RSVP-TE extensions as one of three method recommendations for Distributed Call and Connection Control (G.7713) requirements [4].

However, RSVP's receiver-initiated reservation feature increases the cost in Label Switched Paths (LSPs) establishment, especially in optical networks where there may be a lengthy delay configuring the switching fabric in certain kinds of optical equipment. For example micro mirrors may have to be elevated or moved, and this physical motion and subsequent damping takes time. The overhead reduction in the optical control plane is critical for lightpath modification, rerouting, and fast restoration.

According to the standard procedure description in RSVP-TE [5], the resource reservation is done along the reverse path sequentially, which is inherited from the original RSVP [6]. However, this receiver-oriented resource reservation is time-consuming in optical networks, since optical switches have to take some time to perform physical cross-connections for resource reservation.

This paper proposes an adaptation to do the resource reservation along the forward path. Since an explicit route is usually used in MPLS traffic engineering and calculated by Constraint Shortest Path First (CSPF), the resource requirements are met when CSPF computation is done so that it is unnecessary to delay the resource reservation in the reverse path. Both performance analyses and experimental results show this approach reduces the time cost in lightpath establishment compared with the original scheme and improves performance of the GMPLS control plane. The rest of this paper is organized as follows. Section 2 gives the problem statement of the original RSVP-TE procedure. The proposed adaptation is described in Section 3. Section 4 shows the experimental results based on the real implementation in Linux platform. The paper is concluded in Section 5.

PROBLEM STATEMENT

MPLS is originally proposed to integrate layer 3 routing and layer 2 switching for fast forwarding, however, the most important advantage is traffic engineering support, which has become an extremely important tool for ISPs nowadays as they struggle to keep pace with the ever-increasing volume of Internet traffic. The traditional Interior Gateway Protocols (IGPs) are based on shortest-path first concept, which result in network congestion and traffic load imbalance because they do not take bandwidth availability and traffic characteristics into account when building their forwarding tables. Based on explicit route, traffic engineering is a powerful tool that can balance the traffic load on the various links, routers, and switches in the network so that none of these components is over or under utilized.

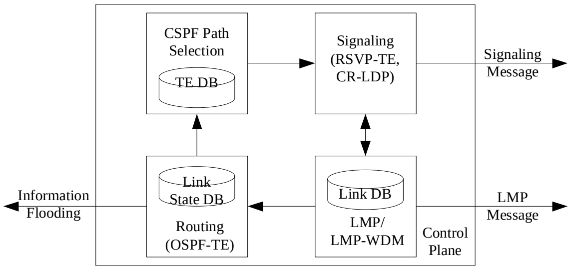

Figure 1. GMPLS Components in the control plane

In GMPLS traffic engineering architecture, there are four main components as shown in the shadowed boxes in Fig. 1, including routing, path selection, signaling, and Link Management Protocol (LMP) components in the control plane. The routing protocols such as OSPF-TE and IS-IS-TE distribute statistic information of network topology as well as dynamic information about network loading, which is specialized in traffic engineering database (TE DB). Each ingress node uses the TE DB to calculate the explicit paths that can be represented by a list of Label Switching Routers (LSRs), which can be routers, switches, Optical Cross-Connects (OXCs) in different domains. Though CSPF computes a path that is thought to be acceptable, the path is not known whether to be able to work until the LSP is actually established by the signaling component. The signaling component is responsible for establishing LSP state, label distribution, and resource reservation. RSVP-TE or CR-LDP is always used as the signaling protocol.

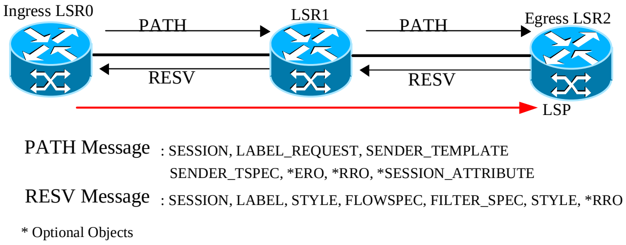

Figure 2. LSP establishment in RSVP-TE

In the standard LSP establishment procedure in RSVP-TE, the PATH message is forwarded from the ingress LSR to the egress LSR going through a number of intermediate LSRs. Once the egress LSR has received the PATH message, it does label allocation and resource reservation. If all those success, the egress node sends the RESV message back to the ingress node. Each intermediate LSR does the same procedure, i.e., label allocation and resource reservation, when it receives the RESV message, then passes it along to its previous LSR. Once the ingress LSR has received the RESV message, the LSP is established successfully. The LSP establishment from LSR0 to LSR2 is shown in Fig. 2.

According to the normal procedure, the traffic control and resource reservation is done on the reverse path, which is inherited from the original receiver-oriented RSVP. This feature has distinguished RSVP from other resource reservation signaling protocols. However, it is not flexible in MPLS-TE and GMPLS context, because it increases time cost of successful resource reservation and degrades the network performance in failed ones.

The resource reservation has to be done sequentially based on the standard description in [6] and [7], which leads to a time-consuming LSP establishment. In packet switched networks, this cost is tolerable, while in lambda and fiber switched networks, i.e., under GMPLS umbrella, the signaling cost is more critical to reduce the control overhead in optical networks.

GMPLS extends MPLS to support suggested label, label set, and bidirectional LSP establishment via one set of signaling protocol messages, which are driven by non-packet switched applications to reduce LSP setup latency and optical device processing time, and to re-

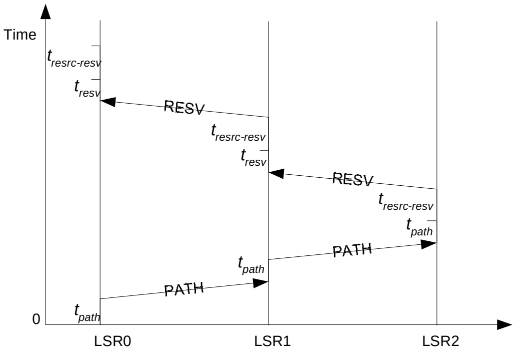

Figure 3. Lightpath setup latency

store traffic to alternate LSPs rapidly in case of network failures [8]. Therefore, the control overhead reduction is important for fast protection and restoration with an acceptable rerouting time of a few milliseconds, e.g., the switching time from a primary path to a secondary path is in less than 50 milliseconds for SONET rings [9]. With Micro Electro-Mechanical system (MEMs) technology, the switching time of light switched from one port to another is under 10 milliseconds for small matrix sizes such as 4×4 or 8×8, while it increases for larger scales. Since optical switches with 1000 × 1000 ports or more are available nowadays, the overhead reduction of resource reservation via RSVP-TE is significant to fast protection and restoration.

PERFORMANCE IMPROVEMENT OF RSVP-TE IN OPTICAL NETWORKS

Processing Modification

Based on the observation and experimentation of our RSVP-TE implementation, we find it possible to minimize the signaling overhead with little modification to the original standard protocol. The basic idea is to move the resource reservation from the reverse path to the forward path without changing the original messages and objects.

In our RSVP-TE design architecture shown in Fig. 4, RSVP-TE signaling protocol is divided into six submodules to handle different functions. Message Processing module deals with message sending and receiving on the network interface. The HELLO message is processed in Neighbor Monitor module to maintain the neighbor relationship, while the other messages such as PATH and RESV are forwarded to Session Control module to create or delete sessions. For each new LSP session, its corresponding information is stored in State Block module. Since PATH, RESV and HELLO messages are required to be refreshed periodically, an independent Timer Manager module is used as timer and counter for sending messages. Moreover, RSVP-TE needs to communicate with other external modules to get necessary information via External Interface.

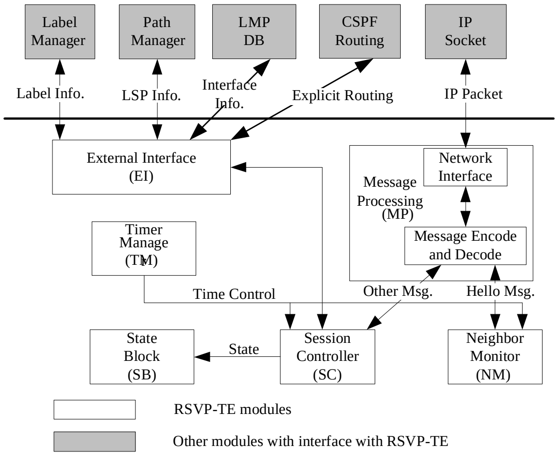

Figure 4. RSVP-TE design architecture

The Session Controller is the core part for processing all kinds of messages except HELLO messages to establish, modify, and delete LSP sessions. The processing flowchart for PATH and RESV messages is illustrated in Fig. 5. Upon receiving RESV message, the resource reservation is done based on the traffic requirements listed in FILTER_SPEC object, which is highlighted in the flowchart. Actually, FILTER_SPEC object is originally composed based on SENDER_TSPEC object when the egress LSR receives PATH message. Therefore, the resource reservation could be done after delivering PATH message. This *forward before reserve* can save reservation overhead, especially in optical network.

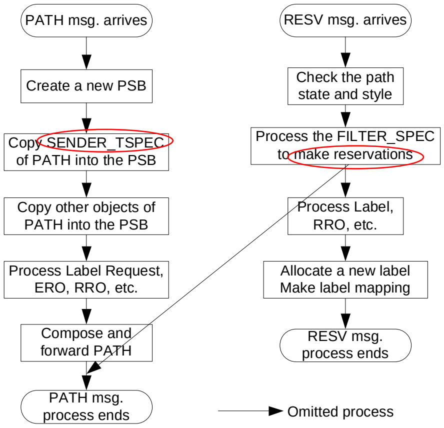

Figure 5: Flowchart for PATH and RESV message processing

In optical networks, the resource reservation and allocation on an OXC include two parts: one is to write the parameters of physical links or ports in database maintained by LMP, which is usually done by software; another one is to select channels and initiate cross-connect fabric, which is done by hardware operations. Once the channels have been selected for a new cross-connection, the remaining time to complete the request consists of the waiting time before the request gets service, and the time to perform the physical cross-connection [10]. The time required to perform the physical cross-connection depends on the technology used in the fabric. While the more worrisome delay is the waiting time before a cross-connection is serviced, and it depends on the cross-connect's ability to process all requests in parallel without any waiting, or one by one with queuing time cost. In the latter case, the waiting time for service degrades performance seriously. Based on our experimentation in RSVP-TE implementation, the most time-consuming part is not the software overhead for resource reservation, e.g., 0.1ms, instead, the physical cross-connect operation usually takes 5ms. Our new adaptation is to send reservation request to cross-connect fabric after processing PATH message in the forward path and check the reply flag indicating the availability of the resource in the reverse path, therefore, the time of resource reservation is overlapped with delivering PATH message on the network.

Performance Analysis

The scheme based on *forward before reserve* in the forward path can reduce not only the time cost of lightpath setup, but also the amount of messages transmitted over the control plane. The basic lightpath setup time consists of the processing time of PATH and RESV messages on each OXC, the time of resource reservation and allocation, and the delay of signaling messages over the control plane. Therefore, the setup time of a lightpath composed of n OXCs can be computed as follows:

$$t_{setup} = t_{path} \times n + t_{resv} \times (n-1) + t_{cross} \times n + 2t_{net} \times (n-1) \tag{1}$$

The processing time of PATH and RESV message in the application layer is denoted by t_{path} and t_{resv} respectively. The resource reservation is to perform the physical cross-connection and the operation time is denoted by t_{cross}. The message transmission time on network is denoted by t_{net}, which actually consists of software overhead of sending and receiving message, and hardware transmission latency on network. Since the egress node sends the RESV message once it has received the PATH message, there is no RESV message processing on this node.

In our new adaptation, the request fro resource reservation is initiated to twist mirrors after sending PATH message, therefore the time to perform physical cross-connections could be overlapped with the network transmission cost. Since the reply from optical cross-connect is checked on the reverse path, this procedure must be done immediately after the PATH message is processed at the egress OXC. Therefore, the time of a lightpath setup with our new scheme is:

$$t_{setup} = t_{path} \times n + t_{resv} \times (n-1) + t_{cross} + 2t_{net} \times (n-1) \tag{2}$$

If an LSP is established successfully, the number of messages is $2(n-1)$ while n denotes the number of nodes in this LSP. This is true to both the original and the new adaptation. However, in the case of resource reservation failure, our new scheme can reduce the amount of messages greatly to avoid network congestion in the control plane. If we assume the resource reservation failed at the xth LSR, therefore, $1 \leq x \leq n$. In the original scheme, the number of PATH messages from the ingress LSR to the egress LSR is $n-1$, while that of RESV, RERR, RTEAR and PTEAR message is $n-x$, $n-x$, $n-1$, and $n-1$ respectively. The total amount of messages to tear down this failed LSP is:

$$n_{msg} = (n-1) + (n-x) + (n-x) + (n-1) + (n-1) \tag{3}$$

Since x is randomly distributed between 1 and n, the average is $\frac{n}{2}$, and the equation (3) can be $4n-3$. However, with our new scheme, the resource is reserved in the forward path, once the failure has been detected, the failed node sends PERR back to the ingress node initiating the PATH message. The ingress node only needs to initiate PTEAR message to the egress node to tear down the unfinished LSP. The average number of messages is $2n-1$ that could reduce almost half amount compared to the original approach.

Addressing Issues Caused by Adaptation

The original RSVP employs receiver-initiated reservation, because the mechanism is based on the IGP hop-by-hop routing and a source can always transmit whether adequate resource exists in the network to deliver the data or not. The receiver knows its own capacity limitations, moreover, it decides which resource should be reserved if network changes in the future. However, in MPLS/GMPLS context, the LSP is usually calculated by CSPF that is constraint oriented. Therefore, the resource in each LSR along the computed LSP meets the requirements and the ingress LSR knows the network resource could satisfy the constraints when it gets the LSP from CSPF, which makes the *forward before reserve* possible.

The receiver-initiated reservation supports Shared Explicitly (SE) style, which is very useful in the resource reservation shared for different LSPs if they have physical links in common. To use which style is determined by the egress LSR and indicated in the style object embedded in RESV message initiated by it. Fortunately, the PATH message has an optional object SESSION_ATTRIBUTE which can indicate the style this session uses. Therefore, the shared resource

Table 1: The lightpath setup time for 3 and 4 OXCs (unit: ms)

Nodes	Scheme	t_{setup}	t_{path}	t_{resv}	t_{cross}	t_{net}
3	Original	**7.04**	0.12	0.063	2.00	0.14
3	New	**3.08**	0.13	0.058	2.00	0.14
4	Original	**9.26**	0.10	0.062	2.00	0.14
4	New	**3.27**	0.11	0.055	2.00	0.14

reservation does not need to wait for the style transmitted by the egress LSR, and it can get adequate information in the forwarding period. The SE style could be supported with the new adaptation

EXPERIMENT RESULTS

The performance of our control plane was investigated using the testbed consisting of four LSRs, which can be regarded as IP routers supporting MPLS or OXCs supporting GMPLS. Since we only investigate the control plane, each node is emulated by a desktop with a 550 MHz Pentium III processor, 256 MB RAM, and running RedHat Linux (2.4.7 kernel based), and 10 Mbps Ethernet interfaces are used between these desktops.

Actually, we implemented RSVP-TE under both Linux and VxWorks operating systems. The experimental results only under Linux platform are shown in this article, because VxWorks saves more software overhead compared to Linux and most of research work in reference are based on Linux. Therefore the experimental results are based on the Linux platform. The PATH message processing time is measured to be 0.1 ms in each node, while that of the RESV message is 0.06 ms. The operating system software overhead of message sending and receiving and the network latency for transmission are measured to be 0.14 ms. Because our evaluation focuses on the control plane without cross-connect data plane, the resource reservation time is set to 2 ms that is to reserve bandwidth and perform cross-connections in our experiments. The experiments results in Table 1 are conformed to the equation (1) and (2). Obviously, the lightpath setup time is proportional to the network size, because the cross-connection operation is time-consuming. While our new adaptation make this cost overlapped with network delay. Therefore, the setup time change little when the network size increases. Because our experiments are based on implementation instead of simulation, the network size is limited to a few OXCs. However, from the performance analysis it is obvious that the performance is better when the network size scales.

CONCLUSION

This article has investigated the time cost of an LSP establishment in RSVP-TE implementation and proposed an efficient adaptation to reserve resource on the forward path since the original reservation on the reverse path is time-consuming in optical networks. Although the receiver-initiated reservation is highlighted in RSVP compared to its previous reservation protocols, this feature is not appropriate in MPLS-TE application, especially for GMPLS, i.e., the physical cross-connect operation cost is high in optical networks. Moreover, to reduce cost, the bi-directional lightpath establishment is an important feature in optical networks, which make the forward path reservation necessary. From the implementation perspective, this article has described the adaptations that reduce the time cost of the original scheme and shown experimental results that the new adaptation is superior to the original scheme when the network size scales. Moreover, it also addresses the implementation problems caused by adaptations.

References

[1] Lixia Zhang, Stephen Deering, Deborah Estrin, Scott Shenker, and Daniel Zappala. RSVP: A New Resource ReSerVation Protocol. *IEEE Network*, pages 8–18, September 1993.

[2] Chuck Semeria. RSVP Signaling Extensions for MPLS Traffic Engineering. *Juniper White Paper, www.juniper.net.*

[3] Chuck Semeria. Multiprotocol Label Switching Enhancing Routing in the New Public Network. *Juniper White Paper, www.juniper.net.*

[4] ITU-T Rec. G.7713.2. DCM Signaling Mechanism Using GMPLS RSVP-TE. February 2003.

[5] D. Awduche *et al.* RSVP-TE: Extensions to RSVP for LSP Tunnels. *IETF RFC 3209*, December 2001.

[6] R. Braden, L. Zhang, S. Berson, S. Herzog, and S. Jamin. Resource ReSerVation Protocol RSVP Version 1, Functional Specification. *IETF RFC 2205*, September 1997.

[7] R. Braden and L. Zhang. Resource ReSerVation Protocol (RSVP) – Version 1 Message Processing Rules. *IETF RFC 2209*, September 1997.

[8] L. Berger. Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description. *IETF RFC 3471*, January 2003.

[9] A. Banerjee, J. Drake, J. Lang, B. Turner, D. Awduche, L. Berger, K. Kompella, and Y. Rekhter. Generalized Multiprotocol Label Switching: An Overview of Signaling Enhancements and Recovery Techniques. *IEEE Communication Magazine*, pages 144–151, July 2001.

[10] Mukul Goyal, Guangzhi Li, and Jennifer Yates. Shared Mesh Restoration: A Simulation Study. In *Optical Fiber Communication Conference*, March 2002.