On the Performance of MPLS TE Queues for QoS Routing

C.J. (Charlie) Liu

AT&T Laboratories

Yihan Li, Shivendra S. Panwar

Dept. of ECE, Polytechnic University
6 MetroTech Center, Brooklyn, NY 11201

I. INTRODUCTION

Traffic engineering (TE) refers to techniques and processes to route traffic through a network on a path other than that would have been chosen if standard routing methods had been used. The goal of traffic engineering to a service provider is to maximize the utilization of network resources, and/or enhance QoS a service provider can offer. To justify the increase in network operational complexity associated with traffic engineering, TE must enable new service offerings, reduce the overall cost of operations, maximize potential revenues and increase customer satisfaction.

In a large network, it is possible that available network bandwidth is not efficiently utilized because the intra-domain routing protocol, such as OSPF, finds path based on a single "least-cost" scalar metric for each destination. This least cost route may not have enough resources to carry all the traffic, or satisfy all the SLA (service level agreement) requirements of passing traffic. Congestion, either for aggregate traffic or for per class traffic, at certain hot spots, can result in sub-optimal use of network resource.

Multi-Protocol Label Switching (MPLS) is an advanced forwarding scheme which extend routing with respect to packet forwarding and path controlling. MPLS TE can relieve congestion for aggregate traffic. In addition, when QoS is deployed in a service provider's network, TE can also help to relieve per class congestion and provide better QoS guarantees.

MPLS TE provides a technique, more elegant and efficient than IP source routing, to allow traffic travel down a path different from conventional IGP destination based hop-by-hop routing. The path is pre-determined at tunnel setup time. Routers along the path do not have to examine the IP header of every passing packet. The basic idea of MPLS involves assigning short fixed length labels to packets inside an MPLS cloud. Throughout the MPLS domains, the labels attached to packets are used to make forwarding decisions. It allows decoupling of the information used for forwarding (a label) from the information carried in the IP header. MPLS TE, using RSVP signaling mechanism [1], injects the notion of a connection to connectionless IP through nailed-up label switched paths (LSP). MPLS TE provides capabilities to specify explicit path for the LSP (ER-LSP) before the LSP is established. We'll refer the nailed-up LSP as MPLS TE tunnel, or simply tunnel, in this paper.

The tunnel explicit routing capability allows routing flexibility. It allows paths, with unequal OSPF cost, to share traffic load. In addition, the Fast Reroute (FRR) feature in MPLS TE allows path restoration within 100 ms in case of link or node failure. In this paper we propose a MPLS TE tunnel mechanism for packet forwarding, which can guarantee the service of real time applications such as VoIP and videoconferencing.

II. USING MPLS TE TUNNEL IN ROUTERS WHEN FORWARDING PACKETS

MPLS TE encompasses the following areas:
(1) *TE information distribution*. OSPF uses Type 10 LSAs [3] to distribute MPLS traffic engineering information.
(2) *Path determination or calculation*. If the path is dynamically calculated, a modified Dijkstra algorithm, often referred to as CSPF (Constrained SPF), at the tunnel head-end is used for this purpose. In CSPF, bandwidth and affinity requirements [4] are among the considerations in addition to the cost of each link. Alternatively, an operator can manually specify the path a MPLS TE tunnel is to traverse. In either case, the tunnel is subject to the same admission control mechanism in path setup time.
(3) *Path setup.* Most vendors choose to use RSVP for path setup.
(4) *Forwarding traffic down a tunnel.*

MPLS TE tunnel is a connection-oriented entity on top of the conventional connectionless IP network. It has been touted by router vendors and IETF activists as a valuable tool to maximize utilization of network resources as described in previous sections. However, MPLS TE admission control mechanism is applied only at the tunnel setup time, not at the packet forwarding time. Bandwidth reservation is policed only at tunnel setup time to limit the number of tunnels traversing a given link. Traffic inside a tunnel has to compete the bandwidth with traffic in other tunnels and regular IP traffic which are not carried by any TE tunnels.

Even though there are benefits of deploying TE tunnels in IP network, there are concerns about its scalability and extra complexity to network operation. For a facility based ISP which owns the physical links and infrastructure of its IP network, the capacity constraint is a relatively minor issue compared to other ISPs which have to purchase or lease capacity from other providers. It is hard to justify sending all IP traffic into fully meshed TE tunnels ubiquitously deployed for a facility based ISP. Instead, only special traffic, such as VoIP or videoconference transported in IP network, are candidates for MPLS TE tunnels. This is different from Diffserv-TE, where all IP traffic are subjected to the same admission control mechanism and have to be carried in TE tunnels to ensure the traffic guarantee.

In this paper we propose to create TE queues for configured MPLS TE tunnels in every router the tunnel traverses, so that traffic that enters a configured MPLS TE tunnel can be preferentially treated by a router's queuing

and congestion avoidance mechanism. A TE queue can be shared by multiple TE tunnels. The TE queue is to be created at tunnel set up time based on the MPLS labels and bandwidth request associated with the tunnel. The bandwidth reserved for each queue is to be set according to the bandwidth of configured tunnels sharing the same queue. The TE admission control mechanism ensures that the sum of the TE queue bandwidth will not exceed the configured RSVP bandwidth of the physical link. The reserved bandwidth can only be used by the traffic carried by the tunnels. At packet forwarding time, the top label in the label stack of each packet carried inside the tunnel will be used as the key for the packet to be sent into the TE queue associated with the tunnel.

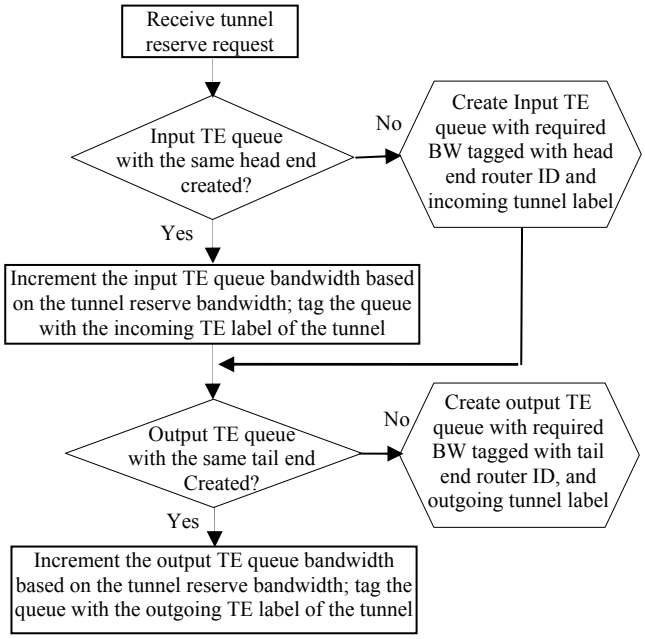

Fig.1 Create Queues for MPLS TE Tunnels

A. TE Queue Creation Process for MPLS TE Tunnels

The proposed MPLS TE Queue creation mechanism at a router is illustrated in the flow chart as shown in Figure 1. We assume both input queues and output queues are implemented in the router. The process below also assumes tunnels with the same head end will share the same input queue, while tunnels with the same tail end will share the same output queue.

During the TE tunnel setup period, the router will query its database to determine if a TE input queue with the same head end had been created. If no such queue had been assigned, the router will create a TE queue, tagged with the head end router ID and the assigned tunnel label, with the requested bandwidth. If a TE queue with the same head end router ID had been created, the bandwidth of the TE queue will be adjusted based on the new tunnel request. The bandwidth adjustment does not have to be the exact increment of the requested bandwidth of the new tunnel. Statistical multiplexing model of tunnel traffic can be incorporated here. The TE queue will also be tagged with one more label, which is the incoming label of the TE tunnel. The output TE queue will be created and setup in a similar fashion.

B. Switching Process for Packets in MPLS TE Tunnels

When a packet is received, a router will determine whether the packet is label switched and whether the label is assigned for a TE tunnel. The packet forwarded via a tunnel will be sent into the appropriate input TE queue based on the incoming label. The router will consult its label-forwarding database to determine its outgoing label. The packet will then be label switched to the appropriate output interface based on the router's scheduling mechanism for input TE queues, and be put into an appropriate output TE queue based on its outgoing label. Then the packet will be forwarded to the next hop based on the router's scheduling mechanism for output TE queues. Because tunnels are envisioned for high priority and demanding traffic only in this paper, it is recommended TE queues assume scheduling priority over all other non-TE queues. Among the TE queues, we assume that the scheduling will be on per packet round robin fashion.

III. ANALYSIS AND SIMULATED PERFORMANCE

A. The system model

The delay a packet suffered from the time it enters the input interface to the time it is transferred to the destined output interface is determined by the scheduling policy of the switch fabric with input TE queues. In this paper we only consider the process from the time packets enter output TE queues to the time they are forwarded to the next hop. We assume that each traffic source can be modeled as a continuous-time Markov process and analyze and simulate the system as a Generalized Processor Sharing (GPS) system [4].

Assume that each input maintains N (output) TE queues and K non-TE queues, as shown in Fig.2(a). All TE queues have the same priority, which is higher than the priorities of non-TE queues. c_n, $1 \le n \le N$, is the guaranteed service rate for TE queue n, and $c = \sum_{n=1}^{N} c_n$. When all TE queues are served, the residual service is distributed to non-TE queues. Each queue n, with the instant rate $r_n(t)$, is modeled as a Markov Modulated Fluid Process (MMFP) with state space $\mathbf{S_n}$, rate matrix $\mathbf{\Lambda_n}$, and infinitesimal generator $\mathbf{M_n}$. The buffer is infinite, and $X_n(t)$ is the occupancy of queue n. For each TE queue, we need to find out the overflow probability with threshold B.

The exact analysis of the system in Fig.2(a) is difficult. To simplify, queue n can be analyzed by a model shown in Fig.2(b) [4]. $r_n^* = r_n(t)$, $r_j^* = \sigma_n r_j'(t)$, and $c_n^* = c_n + \sum_{j \ne n} \sigma_j c_i$, where for *all* $j \ne n$, $\sigma_j = c_n / \sum_{k \ne j} c_k$, and $r_j'(t)$ is the departure process of each queue j obtained while assuming the service rate is c_j. This GPS problem can be resolved with fluid-flow model.

Consider a system as shown in Fig.2(b) but with a service rate c and L independent general MMFP sources. Each source is characterized by $(M^{(i)}, \Lambda^{(i)})$. Then M and Λ of the aggregate source are

$$M = M^{(1)} \oplus M^{(2)} \oplus \ldots \oplus M^{(L)}, \text{and}$$

$$\Lambda = \Lambda^{(1)} \oplus \Lambda^{(2)} \oplus \ldots \oplus \Lambda^{(L)}.$$

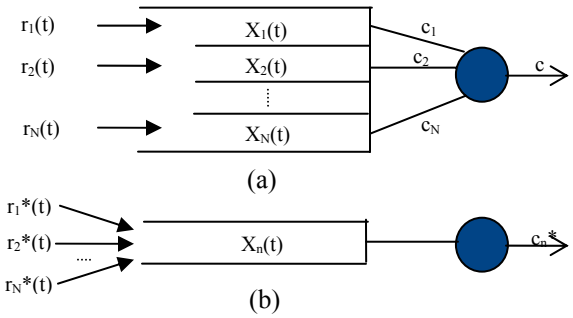

Fig. 2 the analysis model

The Kronecker sum is as follows.

$$A \otimes B = \begin{bmatrix} a_{11}B & a2B & \dots & a_{1n}B \\ a_{21}B & a_{22}B & \dots & a_{2n}B \\ \dots & \dots & \dots & \dots \\ a_{m1}B & a_{m2}B & \dots & a_{mn}B \end{bmatrix}$$

$$A \oplus B = A \otimes I_m + I_n \otimes B$$

where I_m and I_n are the identity matrices of order m and n, respectively. Define state probabilities

$$f_s(x,t) = Prob(S(t) = s, X(t) \le x, t).$$

Then

$$(\Lambda - cI)\frac{df(x)}{dx} = Mf(x).$$

where I is an identity matrix. $(\Lambda\text{-}cI)=D$ is called a drift matrix. If the number of aggregate source state is L, then the result is of the following form

$$f(x) = \sum_{i=1}^{L} a_i \phi_i e^{z_i x}$$

where z_i and I are eigenvalue-eigenvector pairs for the matrix $D^{-1}M$, and a_i are coefficients. a_i corresponding to positive eigenvalues are zero. The number of overload states is the same as the number of negative eigenvalues. $f(x)$ is obtained from the following boundary condition. $f_j(0) = 0$ when $j \in S_0$.

Therefore, the overflow probability, the probability that the buffer occupancy exceeds the threshold B, is as follows.

$$P_{of} = \sum_{j \in S} \Pr ob(S = j, X > B)$$

B. Analysis and simulation results

In this paper we assume that there are three on-off sources, two for TE traffic and one for non-TE traffic. Three cases are considered: (1) all traffic share one queue, (2) all TE traffic share one TE queue and non-TE traffic goes to the non-TE queue, and (3) each TE source traffic goes to its own TE queue and non-TE traffic goes to the non-TE queue. The source parameters are given in Table 1, where α and β are the transition rates from off to on, and on to off, respectively; p is the input rate when the source is on. The guaranteed service rate for source 1 and source 2 are 0.7 and 0.4, respectively, and for source 3 in case 1 is 0.6. The analysis and simulation results of the overflow probability of TE traffic are shown in Fig.3 and 4. As we can see, with the selected system parameters, using TE queue leads to lower overflow probabilities for TE tunnel traffic, and using multiple TE queues can further improve the service of TE tunnel traffic.

Table 1

	α	β	p
Source 1	0.4	1.0	1.2
Source 2	0.4	1.0	1.0
Source 3	1.0	1.0	1.2

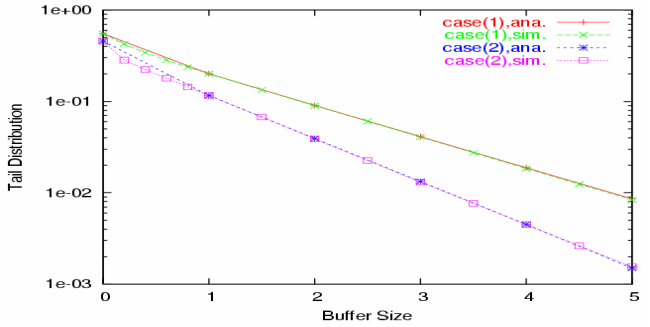

Fig.3 Tail distributions of case 1 and 2.

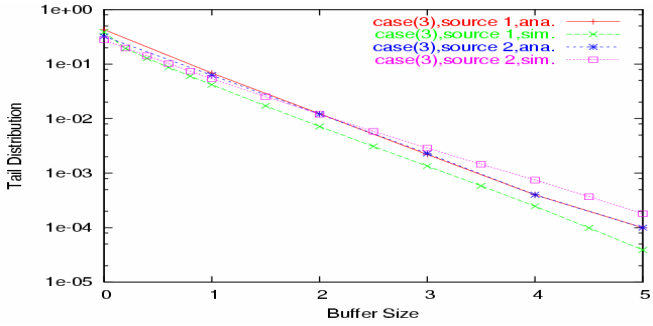

Fig.4 Tail distributions of case 3.

IV. CONCLUSION

The TE queue creation as described in this paper is a new concept to effectively couple the control plane and the data plane for MPLS TE tunnels. The idea takes advantage of the intelligent CSPF routing mechanism for MPLS TE to enable QoS routing. It will save service providers the task and cost of implementing complex bandwidth broker Operation Support System to associate allowed tunnel bandwidth and available queues in the network at the time of provisioning. The mechanism will also ensure an IP network to deliver the stringent QoS required to carry real time traffic such as VoIP. Analysis and simulation results are presented. In our future work, more complicate system model and traffic model will be considered by analysis and simulation.

REFERENCES

[1] [RFC2205] R. Braden, L. Zhang, S. Berson, S. Herzog, and J. Jamin, "Resource Reservation Protocol (RSVP) - Version 1 Functional Specification"

[2] [RFC2370] R. Coltun, "The OSPF Opaque LSA Option"

[3] [RFC2702] D. Awduche, J. Malcolm, J. Agogbua, M. O'Dell, and J. McManus, "Requirements for Traffic Engineering Over MPLS"

[4] S. Mao and S. Panwar, "The Effective Bandwidth of Markov Modulated Fluid Process Sources with a Generalized Processor Sharing Server," Globecom 2001, pp. 2341-2346.