INTER-DOMAIN TRAFFIC ENGINEERING USING MPLS

Meral Shirazipour[1], Samuel Pierre[1], Yves Lemieux[2]

[1] Mobile Computing and Networking Research Laboratory (LARIM),
Department of Computer Engineering, École Polytechnique de Montréal,
C.P. 6079, succ. Centre-Ville, Montréal, Qué., Canada, H3C-3A7
{meral.shirazipour, samuel.pierre}@polymtl.ca
http://www.larim.polymtl.ca/
[2] Ericsson Research Canada
8400 Decarie Blvd., Town of Mount Royal, Québec, Canada, H4P-2N2
{yves.lemieux}@ericsson.ca

Abstract. In the Internet, the traffic crosses between two to eight autonomous systems before reaching its destination. Consequently, end-to-end quality of service requires provisioning across more than one domain. This paper proposes a new scheme for introducing MPLS technology into an inter-domain environment. Results obtained using the OPNET simulation platform show that extending MPLS across AS boundaries can improve the QoS perceived by the end users. This means that inter-domain traffic engineering is a promising solution for a QoS aware Internet.

1 Introduction

For economic reasons, the Internet backbone is the transport network chosen by current and future generation service providers. The foreseen applications have different quality of service (QoS) requirements, in particular regarding availability, delay, jitter, packet loss, etc. It is the job of network operators to assure an adequate support for these applications. They will need to guarantee various levels of QoS to dissimilar traffic flows, while maintaining their own profitability. Numerous architectures have been proposed, or even implemented, in order to increase revenues by maximizing network utilization. But, providing a guaranteed QoS while preserving earnings is a complicated task and remains an open issue.

When speaking of guaranteed QoS, the most important concern is the QoS perceived by the end users. This means that the QoS and application requirements listed above need to be sustained end-to-end, from one user terminal to the other user terminal. It is also known that traffic usually traverses more than one domain in the Internet before reaching its destination (between two to eight domains) [8]. Therefore inter-domain traffic engineering mechanisms are essential in order to achieve end-to-end QoS guaranties.

Providing end-to-end QoS is one of the most important challenges in the Internet. The Internet is composed of about 13000 distinct domains [3], each belonging to a different company or ISP. These domains are under different administrations and are called autonomous systems (AS). An AS is a set of routers functioning under the same administration.

The inter-domain traffic engineering difficulty in the current Internet architecture is caused by the various QoS policies enforced with often a different definition or implementation from one domain to the other. Moreover, topology and link state information is essential for effective inter-domain traffic engineering; but for scalability and privacy reasons the Border Gateway Protocol (BGP) does not propagate this information.

The literature proposes different methods for performing inter-domain traffic engineering [3], [4], [2], [11], [7], [9]. Some of these techniques, e.g. BGP traffic engineering, are already in use in the Internet. Others, such as inter-domain LSP (Label Switched Path) setup or community attribute extensions, are pending proposals at the IETF. BGP traffic engineering is performed by tuning route advertisements. Such tuning mechanisms have their limitations. They are trial and error based, give little control over the end-to-end path taken, lack optimality and have no notion of QoS. The other proposed technique, inter-domain MPLS, is more useful in controlling the traffic, but is not fully defined. It still does not specify how to maintain an end-to-end control over the traffic. This work consists in the description of a method for deploying end-to-end LSPs based on already proposed extensions to RSVP-TE for a similar purpose [9]. In this work, we study the usefulness of end-to-end LSPs by means of simulation results in OPNET. Section 2 gives a brief review on related works regarding the RSVP-TE extensions. Section 3 introduces the proposed mechanism for end-to-end inter-domain LSP setup. Section 4 presents the simulation model and results. Finally, section 5 concludes by presenting related future works.

2 Background and Related Work

Multi-Protocol Label Switching (MPLS) [10] is a packet forwarding framework that performs label switching between layer 2 and layer 3 protocols in the OSI model. The original benefit of MPLS was faster packet forwarding, which nowadays is achievable by more advanced hardware. These days MPLS is mostly used for traffic engineering purposes, to deliver QoS and differentiated services, to offer Fast-Reroute resiliency mechanisms, and to support virtual private networks (VPN). If the MPLS framework is extended beyond a single domain, the technology can be very useful for inter-domain traffic engineering and end-to-end QoS provisioning. The inter-domain extension of MPLS essentially involves the signaling protocol used for the exchange of information between MPLS nodes during LSP setup. Many proposals are made in the literature regarding inter-domain MPLS. Inter-domain LSP setup using bandwidth management points is proposed by [7], but because of its revolutionary nature it is not a pragmatic method in today's Internet. A practicable traffic engineering mechanism designed for the current Internet should allow a smooth migration of the existing technologies towards the proposed one. This can be achieved more easily if the pro-

posed method is an extension to already operating mechanisms. The inter-domain MPLS proposal with RSVP-TE extensions [9] is a method that should consequently be considered. The choice of RSVP-TE in this work can be explained by the fact that it is the most popular and mainly implemented signaling protocol for intra-domain MPLS deployment.

2.1 Intra-domain MPLS versus Inter-domain MPLS

In a MPLS network, Fig. 1, the packet is only routed once at the ingress LER (Label Edge Router) where it receives a label. It is then forwarded through the network following the LSP assigned to its label. At each LSR (Label Switched Router) the label is swapped with another label of local significance (local to the node). When the packet emerges at the egress LER, the last label is removed and the packet is forwarded to its destination with normal IP, or towards its final destination via another network. In each node, packets assigned to a given label belong to the same FEC (Forwarding Equivalence Class). A FEC is a logical entity that designates a group of packets undergoing equivalent forwarding in a given node. During normal IP operation, for each possible next hop, a router usually creates a different FEC. With MPLS, other more advanced criteria can be used to designate a FEC. Criteria such as source-destination address pairs and destination address-ToS pairs are such examples, leaving lots of flexibility for traffic engineering in MPLS networks.

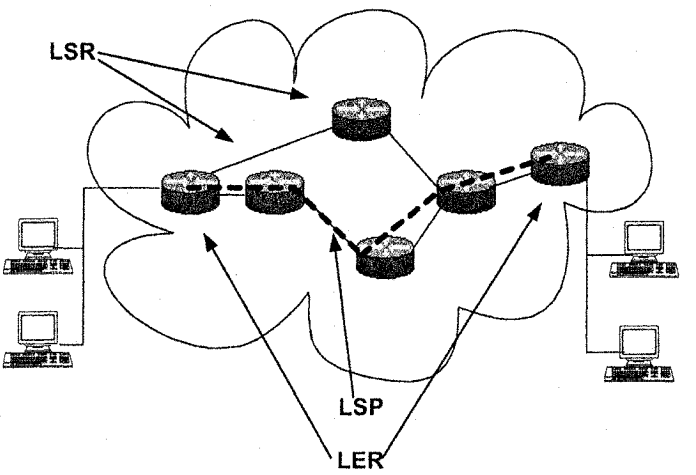

Fig.1: MPLS network

Several requirements for inter-domain MPLS deployment are discussed in [12]. First there is the service providers' need to keep internal resources and intra-domain LSP paths confidential. This implies that global topology information is not available for inter-domain LSP setup. Another requirement for the proposed inter-domain

MPLS mechanism is scalability, in terms of IGP flooding, BGP message exchanges, and signaling extensions. Nonetheless, the usual goals of intra-domain MPLS traffic engineering must be met. These fundamental MPLS goals are

- Support of end-to-end quality of service mechanisms
- Optimization of network resources
- Fast failure recovery methods

2.2 RSVP-TE

In RSVP-TE [1], RSVP [5] is enhanced to enable routers supporting both RSVP and MPLS to associate labels with RSVP flows. To support MPLS, RSVP-TE introduces new objects that will be carried inside RSVP *Path* and *Resv* messages. The LABEL_REQUEST object is carried inside a *Path* message initiated by the LSP's ingress LER. Its purpose is to request the egress LER to initiate a reservation and establish an LSP along the path followed by the *Path* message. The egress LER assigns a label to the LSP that is being created. It then puts that label in the LABEL object of a *Resv* message and sends it to the next node upstream. At each node, a local label is assigned to the LSP, the LABEL object is updated and sent to the next node upstream. This procedure ends at the ingress LER, the LSP being created in this way. RSVP-TE also introduces two other important objects for traffic engineering purposes: the Explicit Route Object (ERO) and the Record Route Object (RRO). These objects are used to allow the LSP to be established along a predefined route rather than the one determined by the IP routing protocol. The predefined route can be calculated by different means, e.g. using manual configuration or by a PCE[1]. The explicitly routed LSP could be used to avoid congested routes, to take disjoint routes during fault recovery mechanisms, or simply to offer the required QoS.

2.3 Extensions to RSVP-TE

In [9], the authors extend the use of RSVP-TE for the deployment of inter-domain LSPs. The goal of the authors is to provide recovery mechanisms for inter-domain link failures, but their method can be refined for the purpose of end-to-end LSP deployment and QoS provisioning. They propose extensions to RSVP-TE that fulfill both the confidentiality and LSP protection requirements. In addition, their method does not disturb the already in place inter-domain routing and signaling protocols. Their LSP establishment method is discussed in the following paragraphs.

For the establishment of intra-domain LSPs, the LER that sets up the LSP tunnel has topology information obtained from the IGP protocols. However, for inter-domain LSPs, the source of the LSP does not have detailed inter-domain topology information. The only information the source router has is on its own domain, obtained by its IGP, along with the route information distributed by BGP. Hence, the

[1] The Path Computation Element concept is being defined at the IETF by the PCE Working Group

source LSR or PCE cannot determine a complete path for the LSP to the destination AS. Moreover, prior to the establishment of the LSP tunnel, the ingress router does not know the IP address of the remote egress router. At this point, the only information the ingress router has is the destination's address prefix and AS number.

To answer this problem, [9] proposes the establishment of inter-domain LSPs based on a destination prefix, or on an AS number and a prefix. For the first case, the LSP is created by forwarding a *Path* message through the network until reaching an LSR with an IP address that matches the prefix. The second case consists again in forwarding a *Path* message on the basis of the destination prefix until reaching an LSR that is part of the specified AS.

It should be noted that the prefix or AS and prefix information are necessary to send the first *Path* message. However upon the reception of the first reservation message, the egress IP address is obtained. It is possible to use this IP address to establish future LSPs to that destination, for backup or load balancing purposes.

Fig. 2 shows an example of the further extension to RSVP-TE that answers the confidentiality requirement of inter-domain MPLS deployment. Here LER R11 of AS1 wants to create an LSP towards AS3. It sends a *Path* message towards that destination but also needs to record in the RRO object the route followed by that LSP. However, recording the complete path of the LSP violates the confidentiality requirement of each AS to keep its internal routing information private. The proposed method [9] consists in aggregating the RRO object in such a way that the only information an AS will divulge about itself is its entry router, its number and its exit router. Table 1 explains Fig. 2 by showing the contents of the RRO object at each point along the LSP. Point 4 demonstrates how the LER R21 of AS 2 is marked as the entry point. Point 7 shows how the exit point of AS2, R23, performs RRO aggregation and hides all the information back to the entry point R21. The source LER, R11, can send subsequent *Path* refresh messages with an ERO containing the path recorded in the RRO. At the entrance of each AS, the ERO will be updated with the RRO of that path-state, containing the information confidential to the AS, which was recorded in the entry LER.

Table 1: *Path* message at different instances in Fig.1

	Dest:	RRO:
①	AS3	R11
②	AS3	R11, R12
③	AS3	AS1, R14
④	AS3	AS1, R14, R21*
⑤	AS3	AS1, R14, R21*, R26
⑥	AS3	AS1, R14, R21*, R26, R22
⑦	AS3	AS1, R14, R21, AS2, R23

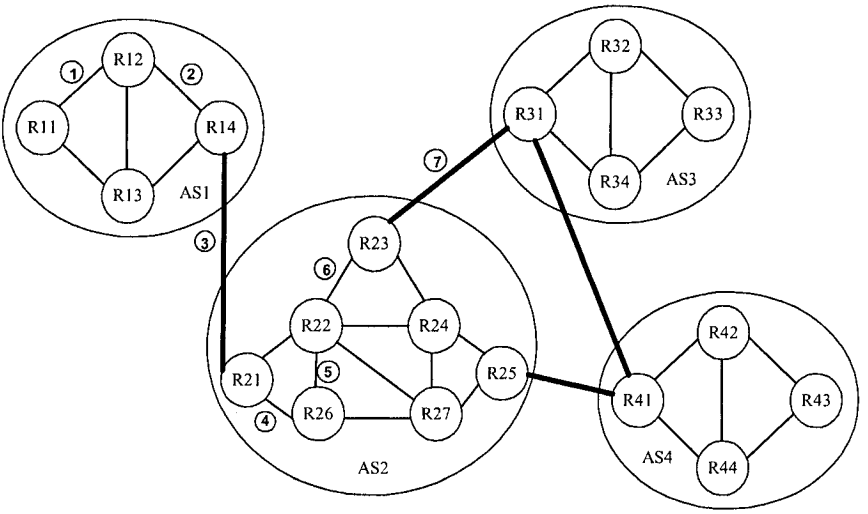

Fig.2: RRO aggregation

3 Proposed Mechanism for End-to-end LSP Setup

To provide an actual end-to-end QoS, the access network which connects the user terminal to the Internet must also be considered. But due to the heterogeneity of access technologies our contribution covers the end-to-end path, up to the last AS, excluding the access network itself. The reader should note that an actual end-to-end mechanism should also cover the access network of the user terminal. Nevertheless, we suggest extending the proposed end-to-end LSP method to the IP operating access networks, in order to support the traffics considered.

To assure the QoS in an inter-domain environment, we use LSPs that cover the end-to-end path of the traffic. The path taken by the LSPs can be optimized using LSP optimization techniques present in the literature [6]. For end-to-end LSPs crossing many domains, this optimization can be performed in a distributed fashion by each AS or by designated PCEs.

For establishing inter-domain LSPs, we make use of the already explained extensions to RSVP-TE [9]. In our suggested mechanism, the head end LER would first give an end-to-end label to the LSP (from the global-LSP subset of its label space), on top of which it would place another label (from the local-LSP subset of its label space). As shown by Fig. 3, at each AS, the packet would be forwarded with a label of minimum depth two [8]. The label at depth 1 will be used to identify the packet as being on an inter-domain LSP. The label at depth 2 will be the actual inter-domain label. In doing so, we intend to propose a differentiation inside the label space used by packets following only the traditional intra-domain LSP and packets following an inter-domain LSP. This differentiation, or packet classification, shall be useful for

future MPLS based inter-domain traffic engineering techniques such as protection, recovery, security, etc.

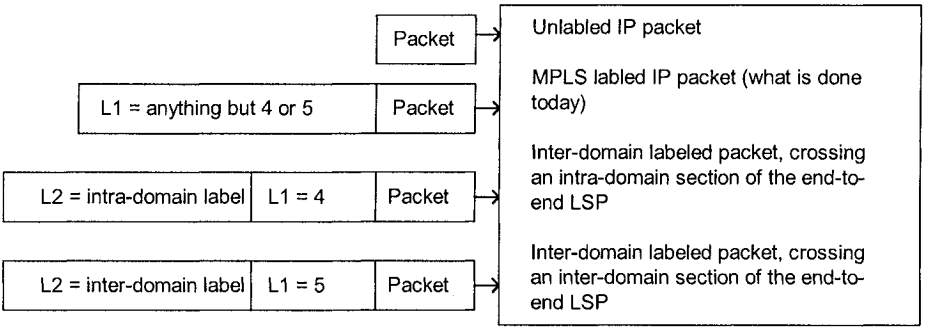

Fig.3: Inter-domain labels (Level 1 and Level 2 labels)

Upon receiving an unlabeled packet, the ingress LER uses the FEC to forward that packet. This decision consist in forwarding with plain IP, i.e. no QoS; along an intra-domain LSP, i.e. leaving part of QoS decisions to further ASs encountered along the way; or through an end-to-end *inter*-domain LSP, i.e. providing end-to-end QoS. If the ingress LER decides to label the packet, it will use the FTN[1] to map the FEC to an NHLFE[2]. Using the information in this NHLFE, it will perform forwarding decisions on the packet. If the incoming packet is already labeled, the ILM[3] will be consulted to forward the packet. In both cases, the labeling will be performed as follows:

- In the case of an LER that does not support inter-domain MPLS, the labeling will be done as described in [10].
- In the other case, the label at level two will be the actual value of the label, while the label at level one will be set to:
  - ✓ 4, if the packet is on the intra-domain part of an inter-domain LSP
  - ✓ 5, if the packet is on the inter-domain part of an inter-domain LSP

4 Simulation Model and Preliminary Results

Our simulation model consists in the implementation of inter-domain LSP scenarios in OPNET modeler 10.5. Fig. 4 shows our network model. This network consists of four ASs. The host applications are located in Montréal, connected to AS 1, and Washington D.C., connected to AS 3. AS 1 can transmit to AS 3 through the path AS 1→AS 2 →AS 3 or through the path AS 1→AS 2→AS 4→AS 3. IGP/BGP routing protocols would normally favor the shortest path, that is AS 1→AS 2 →AS 3. If the

[1] FEC-to-NHLFE, unlabeled packet mapped to an NHLFE [10]

[2] Next Hop Label Forwarding Entry, contains packet's next hop and operations to be performed on label or on packet [10]

[3] Incoming Label MAP, maps incoming labels to NHLFE [10]

AS 2 connection to AS 3 becomes congested or cannot sustain the required QoS anymore, BGP will not reroute the traffic through the AS 2→AS 4→AS 3 path. We demonstrate that by using an inter-domain LSP the traffic can be forwarded through the desired inter-domain route and thus take the less congested one. Moreover, in case of an inter-domain link failure, the LSP fast recovery technique improves the delay experienced by the traffic compared to the time taken by conventional BGP routing to recover from the failure.

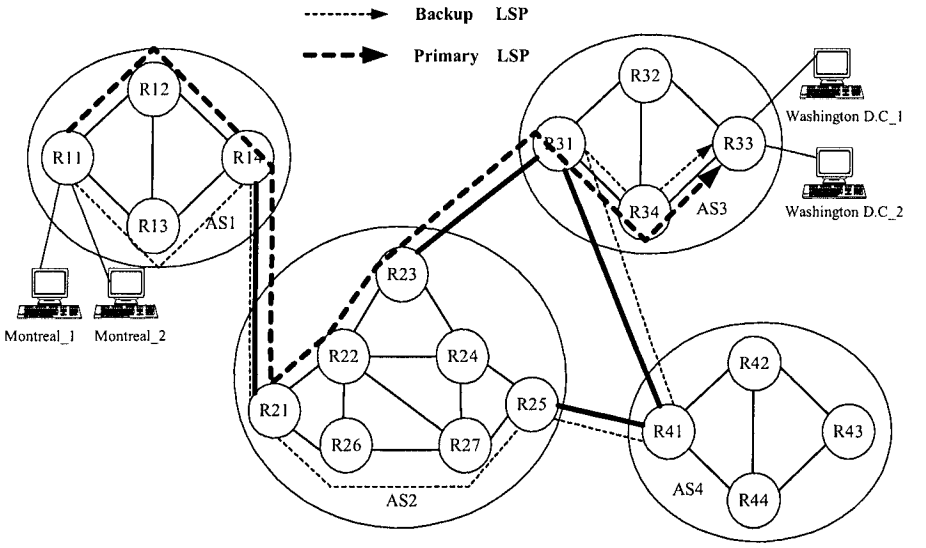

Fig.4: Network model: Internet backbones (ASs)

As depicted in Fig. 4, two LSPs join the source and destination ASs. The following simulation scenarios are sufficient to prove the effectiveness of our scheme compared to normal BGP routing:

Scenario 1: Link R23-R31 becomes congested
Scenario 2: Link R23-R31 fails
Scenario 3: Router R23 fails

In each of the above cases, the traffic is initially forwarded along the R23-R31 link, and after the event of a heavy congestion or a failure, it is switched on the backup LSP in order to avoid the problematic link or node. The simulation results consisted in measuring the QoS gain brought by the use of inter-domain MPLS compared to normal IP-BGP routing. The QoS parameters of interest were the mean delay and the mean jitter experienced by the traffic. Two types of traffic were used: voice traffic and video conferencing traffic. Fig.5 and Fig. 6 show that in the cases of link or node failure, the mean delay and mean jitter experienced were improved for both types of traffic. This is with the exception of the mean jitter experienced by the video conferencing application in the event of a node failure. The difference is insignificant

due to the relatively small values of the jitter. The actual difference between the mean jitter with our scheme compared to BGP routing is 0.00004 s. In the case of congestion avoidance, it is seen that the gain is negative for both types of traffic and for both QoS parameters. The reason for that is that the congestion level of the link was not acute. Since our backup LSP takes a longer route to the destination, the delay and jitter experience with MPLS were slightly higher. An acute congestion scenario is equivalent to our failure scenarios. What is interesting with the use of inter-domain LSPs is the predictability it brings in the QoS experienced by the traffic. It is a method for load balancing in the Internet. Moreover, knowing in advance the path (LSP) taken by the traffic will permit us to estimate bounds on the QoS parameters of interest.

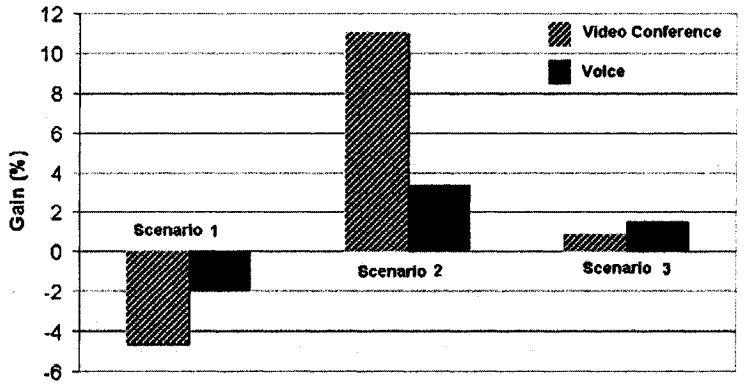

Fig.5: Gain in mean delay with inter-domain MPLS

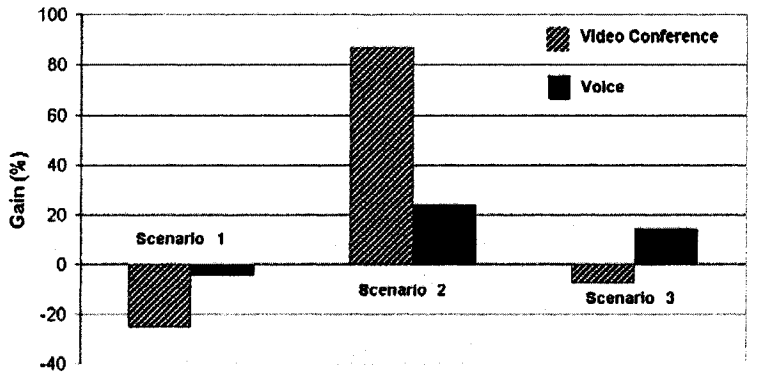

Fig.6: Gain in mean jitter with inter-domain MPLS

5 Conclusions

This paper consisted in refining and applying the already proposed extensions to RSVP-TE, to deploy MPLS across AS boundaries in order to achieve end-to-end control over the traffic in the Internet. Since Internet traffic crosses a few AS boundaries before reaching its destination, providing end-to-end QoS necessitates achieving end-to-end control of the traffic. MPLS is one of the best solutions for end-to-end control of the traffic and for end-to-end QoS provisioning, since it already serves these purposes inside ASs. Our future objectives are to define the signaling of the required QoS and to propose an end-to-end QoS provisioning architecture in the Internet.

References

1. D. Awduche, L. Berger, T. Li, V. Srinivasan, G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, Dec. 2001.
2. O. Bonaventure, S. De Cnodder, B. Quoitin, R. White, "Controlling the redistribution of BGP routes", April 2003. Work in progress, draft-ietf-grow-bgp-redistribution-00.txt.
3. O. Bonaventure, B. Quoitin, S. Uhlig, C. Pelsser, L. Swinnen, "Interdomain Traffic Engineering with BGP", *IEEE Communications magazine*, Vol. 41, No. 5, May 2003, pp.122-128.
4. O. Bonaventure, B. Quoitin, "Common utilization of the BGP community attribute", June 2003. Work in progress, draft-bonaventure-quoitin-bgp-communities-00.txt.
5. R. Braden, L. Zhang, S. Berson, S. Herzog, S. Jamin, "Resource ReSerVation Protocol (RSVP)", RFC 2205, Sept. 1997.
6. M. Girish, B. Zhou, J.Q. Hu, "Formulation of the Traffic Engineering Problems in MPLS based IP Networks," *Proceedings of the Fifth IEEE Symposium on Computers and Communications (ISCC 2000)*, Antibes, France, July 4-6, 2000, pp. 214-219.
7. T. Okumus, J. Hwang, H.A. Mantar, S.J. Chapin, "Inter-Domain LSP Setup Using Bandwidth Management Points", *Proc. of IEEE Global Communications Conference Globecomm 2001*, Nov. 2001, San Antonio, TX, USA, pp.7-11.
8. P. Pan, "Scalable Resource Reservation Signaling in the Internet", PhD thesis, Columbia University, 2002, 164 pages.
9. C. Pelsser, O. Bonaventure, "Extending RSVP-TE to support Inter-AS LSPs", *2003 Workshop on High Performance Switching and Routing (HPSR 2003)*, June 24-27th, 2003, pp.79-84.
10. E. Rosen, A. Viswanathan, R. Callon, "Multi-protocol Label Switching Architecture", RFC 3031, Jan. 2001.
11. S.R. Sangli, D. Tappan, Y. Rekhter, "BGP Extended Communities Attribute", Aug. 2003. Work in progress, draft-ietf-idr-bgp-ext-communities-06.txt.
12. R. Zhang, J.P. Vasseur, "MPLS Inter-AS traffic engineering requirements", May 2003. Work in progress, draft-zhang-mpls-interas-te-req-03.txt.