Extending RSVP-TE to support Inter-AS LSPs

Cristel Pelsser
University of Namur (Belgium)
Email : cpe@info.fundp.ac.be

Olivier Bonaventure
Université Catholique de Louvain (Belgium)
Email: Bonaventure@info.ucl.ac.be

***Abstract*— Multiprotocol Label Switching (MPLS) is currently used inside Autonomous Systems (ASs) for Virtual Private Networks (VPNs) or Traffic Engineering purposes. We first discuss the Service Provider's requirements for the utilization of MPLS Label Switched Paths (LSPs) across AS boundaries. Then we propose a minimum set of extensions to RSVP-TE that allow to establish inter-AS LSPs in accordance with the SPs' requirements. We also show how LSP protection techniques can be extended to provide link or node failures protection for the inter-AS links and border routers.**

I. Introduction

Multiprotocol Label Switching (MPLS) is currently mainly used inside ASs, also called "domains", to provide Virtual Private Networks (VPNs) services or for traffic engineering or fast restoration purposes. Several fast restoration techniques have been proposed [10], [11]. They rely on failure notifications, pre- or on-demand establishment of backup paths and switching traffic to the backup paths when a failure notification is received, as exposed in [15].

Besides their utilization inside ASs, the use of MPLS and GMPLS across AS boundaries could be an efficient solution to support inter-AS VPNs [14], to build more scalable Internet eXchange (IX) points [8] or to provide shorter restoration times than those obtained with the Border Gateway Protocol (BGP) in case of inter-domain failures[5], [6].

The support of inter-AS VPNs becomes more and more important as customer's VPNs span over multiple Service Providers (SPs) [18]. In the meantime, the customer still requires to maintain a set of performance targets, in terms of bandwidth, delay, and/or delay jitter for VPN traffic [18] which cannot be ensured when a VPN is established by relying only on the Border Gateway Protocol (BGP) as proposed in [14].

Current IXs are often based on switched LANs or ATM switches. This creates several problems, for example, bandwidth limitation, operational cost, low scalability, and dependency on data-link mediums as shown in [8]. An interesting architecture, proposed in [8], would be to base new Internet eXchange points on MPLS. Those IXs would require mechanisms to establish inter-AS LSPs.

When an inter-domain failure occurs, BGP may take several minutes to reach a consistent view of the network topology after the fault [5], [6]. This long restoration time of BGP is clearly a problem when using the Internet for mission critical services. The use of MPLS for inter-AS traffic forwarding would provide better restoration times than BGP because LSPs could be protected against link, node, segment failures, and could be established on-demand.

A. Inter-AS LSP requirements

In [18], several requirements for MPLS Inter-AS Traffic Engineering (TE) are expressed. Among these requirements is the desire of SPs to keep internal AS resources and the set of hops followed by the TE-LSP confidential. This confidentiality requirement implies the capability of partly specifying the hops that the TE-LSP must traverse since global topology information is not available. Moreover, it must be possible to perform path optimization inside each transited AS, where the required information is available.

A second requirement concerns the restoration capabilities of inter-AS LSPs. The proposed solution has to be able to provide rapid local protection against link and node failures. Additionally, it should support the establishment of multiple link/Shared Risk Link Group (SRLG)[1]/node diversely routed inter-AS TE LSPs between a pair of Label Switching Routers (LSRs).

A last requirement is that the proposed solution should be scalable in terms of the amount of IGP flooding, the additional information carried by BGP and the amount of signaling messages exchanged.

In the first part of the paper, we show how intra-AS TE LSPs are established and protected against link and node failures. Then, we present our solution for the establishment and the protection of traffic engineered inter-AS LSPs. Finally, we compare our proposal with other solutions to the inter-AS LSP establishment problem.

II. Intra-AS LSPs

The specification of RSVP-TE [1] defines extensions to the Resource reSerVation Protocol (RSVP) in order to establish traffic engineered LSPs. Among these extensions are the ability to distribute labels and to specify a strict or a loose path to be followed by an LSP.

RSVP messages are composed of an header followed by a sequence of objects. Among these objects are the Session Object, the Sender Template Object, the Explicit Route Object (ERO) and the Record Route Object (RRO). The Session and the Sender Template Objects are used to identify an LSP.

[1] An SRLG identifies a set of links that may fail together. If one of the links belonging to an SRLG fails, all the other links belonging to the same SRLG may also be impacted by the failure.

Based on the values stored inside these objects, a router is able to create and/or access the path-state related to this LSP.

The routing of RSVP-TE `Path` messages is performed on the basis of the ERO, when it is present. This object contains a list of subobjects representing abstract nodes to be crossed by the LSP. Abstract nodes may either be a single node or a group of nodes such as a network prefix or even an entire AS. Subobjects inside the ERO can be marked with a "loose bit" to indicate that the subobject may be reached after crossing nodes that are not present inside the ERO[2]. Intermediate LSRs may complete the ERO when they meet an abstract node or a node marked with the "loose bit" inside the ERO. When no ERO is present inside a `Path` message, it is routed as a normal IP packet based on the packet's destination, i.e. the IP Destination Address (IPDA).

The path of an LSP can be recorded by using the RRO. This object is inserted inside `Path` and `Resv` messages by their source. Each LSR crossed by such message adds its address inside the RRO and stores the RRO inside the LSP's path-state. By inserting the RRO both inside `Path` and `Resv` messages, each LSR on the path of the LSP can obtain the complete path of the LSP. This information is useful for loop detection, route pinning and for the computation of disjoint LSPs.

A. Protection of intra-AS LSPs

There are different ways to protect intra-AS LSPs [15]. A primary LSP may be end-to-end protected with a secondary LSP, disjoint from the primary LSP, joining the same head-end[3] and tail-end [4] LSRs. An alternative is to protect segments of the primary LSP with other LSPs that are disjoint from the primary LSP's segments. This is called "local protection" and it enables to protect an LSP against single link or node failures.

In case of failure, a message notifying the failure has to travel all the way back to the head-end LSR when the LSP is end-to-end protected. On the other hand, if the LSP is segment protected, the notification message only travels backward to the Point of Local Repair (PLR), that is the head-end of the backup LSP. This router is close to the point of failure allowing faster restoration times than with end-to-end protection.

Local protection can be provided by Detour LSPs or Bypass Tunnels. We shortly describe the establishment of Detour LSPs [11]. For the use of Bypass Tunnels[5], we refer the reader to [11].

A Detour LSP protects against a node failure and against the failure of the link used by the primary LSP to join that node, i.e. its upstream link. In order to protect an LSP against single link and node failures with a Detour LSP, two objects are required: the FAST_REROUTE Object and the DETOUR Object. The Fast Reroute Object is carried inside the `Path` message of the primary LSP and indicates the type of protection required by the primary LSP. The Detour Object is carried inside the `Path` message of the Detour LSP, i.e. the LSP protecting a segment of the primary LSP. The Detour Object contains the address of the PLR and of the node to be avoided.

The node, where merging of the Detour LSP with the primary LSP occurs, is called the "Path Merge LSR (PML)". This LSR may be any router on the path of the primary LSP downstream from the PLR and the node to protect.

III. Inter-AS LSPs

In this section, we describe our solution to establish inter-AS LSPs. We introduce the RSVP-TE extensions required to enforce the requirements presented in section I-A. A detailed description of these extensions may be found in [12].

The desire of SPs to hide their internal topology, as currently achieved by BGP and the need for LSP's protection are not easily satisfiable simultaneously. Indeed, it is necessary for an LSR to know the path of an LSP to be able to protect it. This information is easily obtained from RSVP-TE objects for the intra-AS path of an LSP (see II) but it is not so obvious to obtain such information for its inter-AS path when the confidentiality requirement regarding internal AS's topologies is observed. In this paper, we propose extensions to RSVP-TE that fulfill both the confidentiality and the protection requirements concurrently while trying to keep our solution scalable. Our solution also tries to only impact the head-end LSR, the intermediate AS Border Routers (ASBRs) on the path of the inter-AS LSP and the tail-end LSR of the LSP therefore allowing a smooth migration toward the support of inter-AS LSPs. Our solution does not impact the current BGP and MPLS Traffic Engineering techniques. Moreover, it does not require additional IGP flooding. And last but not least, our solution supports the dynamic establishment of inter-AS LSPs avoiding the need for static configuration at the head-end LSR of the inter-AS LSP.

A. Destination of an LSP

The first problem encountered in the dynamic establishment of inter-AS LSPs is that unless the head-end LSR has been manually configured with the IP address of the tail-end LSR, it cannot obtain this information, before establishing the tunnel, on the basis of its BGP routing table, which contains only information about destination prefixes and their AS paths.

To solve this problem, we propose to enable the establishment of LSPs based on a prefix or on an AS number and a prefix destination. During the establishment of an LSP based on a prefix destination, the `Path` message is forwarded through the network until it reaches an LSR with an IP address that belongs to this prefix . The `Path` message itself is routed on the basis of its destination IP prefix and possibly along an explicit route defined by an Explicit Route Object (ERO). The second type of destination that we propose is composed of an AS number and an IP prefix. In this case, the `Path` message is forwarded through the network on the basis of the destination

[2] When the "loose bit" is not set, the `Path` message has to reach the following node inside the ERO without crossing intermediate nodes.

[3] The head-end LSR is the router that initiated an LSP and is the source of this LSP.

[4] The tail-end LSR is the last router on the path of an LSP; it is the destination of this LSP.

[5] A Bypass Tunnel is an LSP that protects a set of LSPs crossing common resources.

prefix until it reaches an LSR that is part of the specified AS independently of the destination prefix. The path followed by the `Path` message can also optionally be specified with an ERO object. It is necessary to specify a prefix in addition to the AS number because BGP only provides prefix based routing.

These AS+prefix or prefix destination types are necessary to send the first `Path` message. However, once the first `Resv` message has been received, the source LSR of the LSP knows the IP address of the destination LSR. But, since the identification of an LSP is composed of the destination of this LSP. It is not desirable to change this destination once the LSP has been established and, therefore, the same destination is used for consecutive `Path` refresh messages.

B. Explicit routing of an LSP

The Explicit Route Object (ERO) is well suited for the establishment of inter-AS LSPs in that it permits the head-end of the LSP to partially compute the path to be followed by the LSP. Following nodes crossed by the `Path` message are able to complete this object as the `Path` message goes along. More precisely, the head-end LSR is only able to fill the ERO with nodes that belong to the same AS and eventually with the ASs that will be crossed by the `Path` message. At the entrance of each AS, the ASBR computes the path of the LSP toward the downstream AS and completes the ERO accordingly. This is illustrated in figure 1 where $R0$ computes the path toward $AS1$ and sets the ERO accordingly. Inside $AS1$, $R3$ completes the ERO toward the next AS, $AS3$ and so on. These paths are computed based on the destination prefix: $65.0.0.0/8$.

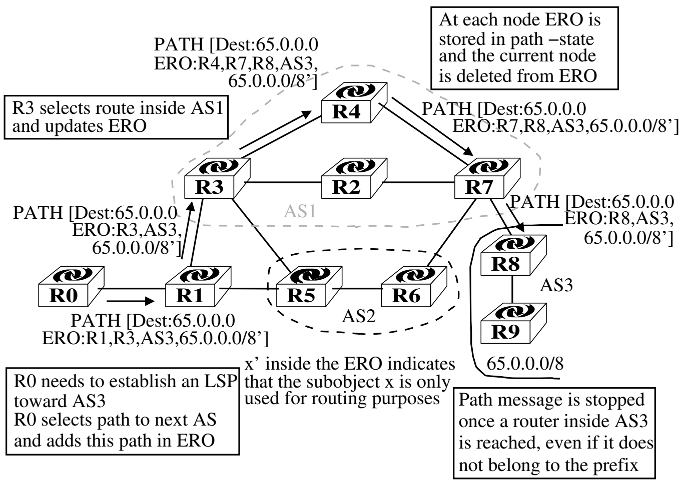

Fig. 1. Establishment of an inter-AS LSP

The ERO object may be constructed at the head-end LSR either based on a manual configuration that specifies the ASs and/or the ASBRs to be crossed by the LSP, based on the BGP routing table or based on a Path Computation Server (PCS)[6]. The inter-domain path selection could be performed by relying on QoS information distributed by extensions to BGP proposed in [17] and [4]. The ERO specifies only a set of hops on the path of the inter-AS LSP and it leaves each crossed AS the responsibility of the local path optimization according to a set of constraints also carried inside the `Path` message of the LSP. This fulfills requirements from the first paragraph of section I-A.

C. RRO aggregation

The Record Route Object (RRO) enables to obtain the path followed by an LSP leading to its usefulness in detecting loops inside the LSP's path, the capability to pin the LSP onto its path and the possibility to compute LSPs disjoint from this LSP for end-to-end or local protection.

We note that recording the path of an inter-AS LSP may be in contradiction to the SPs desire to hide the internal topology of ASs. Therefore, we propose to modify the processing of this object at the ASBRs so as to withhold from neighboring ASs the complete path followed by the LSP inside the current AS. We call this process "RRO aggregation".

The aggregation of the RRO consists in marking the subobject added by the entry ASBR inside the AS. And, at the last router of the AS, i.e. the exit ASBR, the subobjects starting from the marked subobject, added by the nodes inside the AS, are removed. These subobjects are replaced by the address of the entry ASBR, the AS number and the address of the exit ASBR in order to keep enough information to perform loop detection, disjoint path computation and route pinning of the inter-AS LSP. Figure 2 illustrates the aggregation of the RRO. In this figure, $R3$ adds its address inside the RRO and marks it. The following LSRs ($R4$ and $R7$) add their address inside the RRO. The exit ASBR, $R7$ in figure 2, removes all addresses starting from the marked subobject, representing the address of $R3$. It replaces these subobjects by the address of the entry ASBR ($R3$), its AS number ($AS1$) and its own address ($R7$).

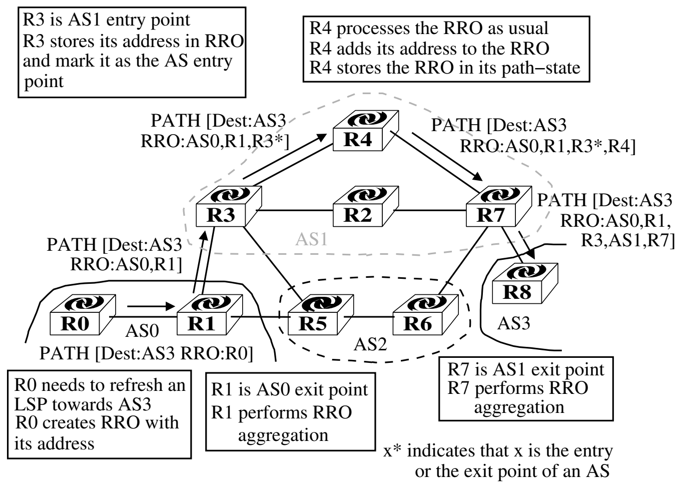

Fig. 2. Processing of the RRO object

The modification of the RRO processing that we propose, only takes place at ASBRs and gives the opportunity to hide

[6]A PCS is a path computation tool with whom LSRs may communicate with RSVP path computation request and reply messages as defined in [16].

the internal topologies of ASs while still permitting to protect the established inter-AS LSPs, in conformance to the SPs' requirements.

D. Protection of inter-AS LSPs

In this section, we look at the establishment of LSPs that are totally or segment disjoint from an existing inter-AS LSP. The first objective is to provide restoration capabilities analogous to the ones provided to intra-AS LSPs including local protection against link, node and SRLG failures. Further, the possibility to establish completely link or node disjoint LSPs can be useful to balance traffic on these disjoint LSPs and may be used for end-to-end protection.

As proposed in the previous section, the RRO records the aggregated path of an LSP, which is necessary for the computation of disjoint LSP segments. It informs each LSR, on the path of the LSP, about the ASs, the entry and the exit ASBRs crossed by the LSP in addition to the complete path of the LSP inside the AS. Based on this information, different types of protection may be provided to an inter-AS LSP but we favor segment protection over end-to-end protection of LSPs in order to leave local operators flexibility in the choice of their protection policy and to achieve faster traffic recovery. For a description of the establishment of end-to-end links or nodes disjoint LSPs for link/node protection or load balancing purposes we refer the reader to [12].

Techniques to protect AS core nodes and links joining these nodes are described in [11]. Here, we only consider the protection of ASBRs and of their upstream link, due to space limitations. The protection of links belonging to distinct ASs, called "inter-domain" links, is discussed in [12]. These techniques can be combined with the ones described in [11] to protect inter-AS LSPs all the way along their path.

Figure 3 shows a reference configuration and the information required at the different routers in order to protect, with a Detour LSP, a primary LSP against the failure of an exit ASBR ($R13$ on figure 3) and its upstream link on the LSP's path ($R11 - R13$).

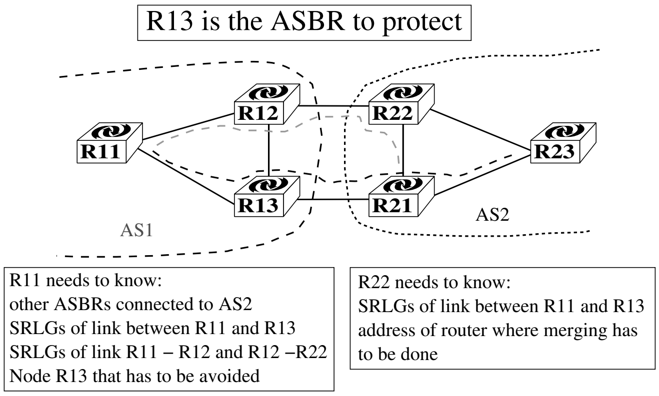

Fig. 3. Node protection of the exit ASBR with a Detour LSP

To protect an exit ASBR ($R13$ on figure 3), the LSR upstream of the exit ASBR, the PLR ($R11$), needs to be able to determine the path for the Detour LSP. Therefore, the PLR needs to find another ASBR inside its AS that is also connected with the downstream AS ($AS2$). This information can be obtained through manual configuration or distributed by iBGP if the PLR receives routes via iBGP. If the PLR does not receive BGP routes, then it should communicate with another LSR to obtain the required information. We propose to do this via a dedicated Path Computation Server (PCS) or by using the PCS protocol, proposed in [16], to contact the exit ASBR to be protected ($R13$).

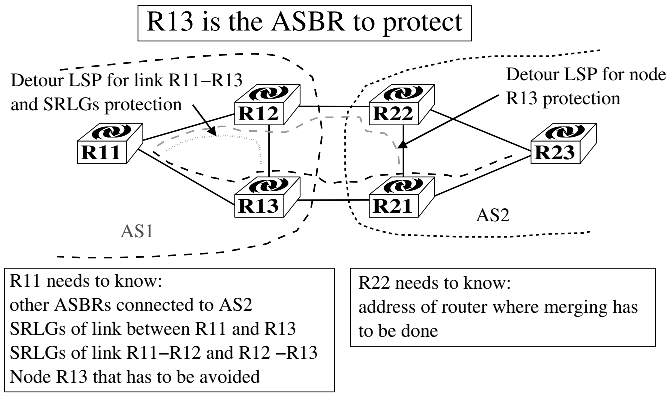

Fig. 4. Node protection of the exit ASBR with Detour LSPs

If the Detour LSP also has to be SRLG disjoint in addition to being link and node disjoint, two Detour LSPs need to be established, as shown in figure 4. The first Detour LSP protects against the failure of the link ($R11 - R13$) and its SRLGs while the other Detour LSP protects against the node failure ($R13$) [3]. Furthermore, the Detour LSP protecting the SRLGs must merge with the primary LSP inside the AS containing the link to protect ($AS1$). The utilization of two LSPs is necessary because neighboring ASs use different SRLG numbering schemes. For example, links $R11 - R13$ and $R22 - R21$ may share the same physical infrastructure but be assigned different SRLG values inside $AS1$ and $AS2$. Therefore, it is not possible for $AS2$ to determine the SRLGs from which the Detour LSP has to be disjoint in order to provide SRLG protection[7]. The second Detour LSP avoids the exit ASBR $R13$ and merges with the primary LSP inside the downstream AS.

Since merging of the Detour LSP and the primary LSP at the PML depends on the AS local policy, such as, for example, merging at the nearest node based on the IGP metric, we consider the use of a PCS to obtain the address of the PML and the path to reach this PML, when the PML is not located in the same AS as the PLR. The PCS may be a dedicated server or the entry ASBR of the primary LSP. The entry ASBR of the Detour LSP ($R22$), contacts this PCS and obtains the information on the detailed path of the primary LSP inside the AS that it is missing locally.

[7] In addition, $AS2$ doesn't even know the existence of link $R11 - R13$ since $AS1$'s inside topology is hidden from $AS2$

In order to protect an inter-AS LSP against the failure of an entry ASBR ($R21$ on figure 3), the same type of information is required by the PLR ($R13$). In this case, the PLR is the exit ASBR upstream from the entry ASBR to protect and therefore, it runs BGP and can obtain information concerning alternative inter-domain links from its Routing Information Base (RIB). It does not need to communicate with a PCS to obtain these links. Therefore, the resulting solution is simpler than for the protection of exit ASBRs and may be found in [12].

To protect against the SRLG failure of the inter-domain link ($R13 - R21$), the PCS or the entry ASBR of the primary LSP needs to know these SRLGs in order to compute the disjoint path. Therefore, we suggest, as proposed in [3], to store inside a new object, defined in [7], a reference to the link whose SRLGs are to be avoided. The entry ASBR of the Detour LSP communicates this object to the PCS. This server can then obtain the SRLGs of the link, inside this AS, and compute a disjoint path toward the PML.

The establishment of Bypass Tunnels for the protection of inter-AS LSPs is analogous to the establishment of Detour LSPs, exposed previously. However, the selection of an already established Bypass Tunnel requires additional mechanisms. An overview of these mechanisms may be found in [12].

IV. Related Work

Few papers have discussed solutions to allow the establishment of LSPs across AS boundaries. In [9], a solution based on the utilization of a specialized Bandwidth Broker agent relying on the SIBBS inter-domain signaling protocol is proposed. Our solution based on RSVP-TE has several advantages over the utilization of a special inter-AS signaling protocol. First, RSVP-TE is already implemented and deployed, which is not the case of SIBBS. Second, our extensions could be added to existing RSVP-TE implementations with a limited amount of effort.

Another solution is the utilization of the BGP extension defined in [13] to distribute MPLS labels and thus establish inter-AS LSPs. Compared with our solution, a drawback of this BGP approach is that with BGP, the inter-AS LSPs are established without being able to specify bandwidth or fast restoration constraints.

[2] proposes two BGP extensions to allow the establishment of optical inter-domain paths. The first extension allows to distribute reachability information by defining a new BGP multiprotocol extension and using extended communities to encode lightpath information. The second extension proposes to use BGP to setup inter-domain lightpaths. This setup is based on the utilization of a BGP update message containing special extended communities. This second extension has several drawbacks compared to our solution. First, [2] only addresses the signaling of the lightpath between domains, it does not discuss how an inter-domain path should be established inside each transit domain while our solution works both inside and outside domains. Second, the BGP extensions described in [2] do not allow to specify fast restoration or QoS requirements for the path being established.

V. Conclusion

Although MPLS and GMPLS are currently used only inside ASs, applications such as inter-AS VPNs, inter-AS fast-restoration and traffic engineering force network operators to also consider those technologies across AS boundaries. In this paper, we have first discussed the requirements for the establishment of such inter-AS LSPs. We have then shown that by introducing a limited number of protocol extensions, it is possible to establish inter-AS LSPs with local protection while still preserving the confidentiality requirement of network operators. Our protocol extensions, described in more details in [12], support Bypass tunnels, Detour LSPs and also allow to establish disjoint inter-AS LSPs for load balancing or end-to-end restoration.

Acknowledgments

This work was supported by the European Commission within the IST ATRIUM project. We would like to thank Stefaan De Cnodder and Louis Swinnen for their comments.

References

[1] D. Awduche, L. Berger, D.-H. Gan, T. Li, V. Srinivasan, and G. Swallow. RSVP-TE: Extensions to RSVP for LSP Tunnels, December 2001. RFC 3209.

[2] M. Blanchet, F. Parent, and B. St-Arnaud. Optical BGP (OBGP): InterAS lightpath provisioning, March 2001. Work in progress, draft-parent-obgp-01.txt.

[3] S. De Cnodder and C. Pelsser. Protection for inter-AS MPLS tunnels, February 2003. Work in progress, draft-decnodder-mpls-interas-protection-00.txt.

[4] G. Cristallo and C. Jacquenet. Providing quality of service indication by the BGP-4 protocol: the QOS_NLRI attribute, July 2001. Work in progress, draft-jacquenet-qos-nlri-03.txt.

[5] C. Labovitz, A. Ahuja, A. Bose, and F. Jahanian. Delayed internet routing convergence. In *SIGCOMM*, pages 175–187, 2000.

[6] C. Labovitz, A. Ahuja, R. Wattenhofer, and V. Srinivasan. The impact of internet policy and topology on delayed routing convergence. In *INFOCOM*, pages 537–546, 2001.

[7] C.-Y. Lee, A. Farrel, and S. De Cnodder. Exclude routes - extension to RSVP-TE, March 2003. Work in progress, draft-lee-ccamp-rsvp-te-exclude-route-02.txt.

[8] I. Nakagawa, H. Esaki, and K. Nagami. A design of a next generation IX using MPLS technology. In *2002 Symposium on Applications and the Internet (SAINT'02, IEEE)*, Nara, Japan, Jan 28th-Feb 1st 2002.

[9] I. T. Okumus, J. Hwang, H. A. Mantar, and S. J. Chapin. Inter-domain LSP setup using bandwidth management points. In *IEEE Globecom*, San Antonio, Texas, 25-29 November 2001.

[10] K. Owens, V. Sharma, S. Makam, C. Huang, and B. Akyol. Extensions to RSVP-TE for MPLS Path Protection, July 2001. Work in progress, draft-chang-mpls-rsvpte-path-protection-ext-02.txt.

[11] P. Pan, D.-H. Gan, G. Swallow, J.-P. Vasseur, D. Cooper, A. Atlas, and M. Jork. Fast reroute extensions to RSVP-TE for LSP tunnels, January 2002. Work in progress, draft-ietf-mpls-rsvp-lsp-fastreroute-00.txt.

[12] C. Pelsser and O. Bonaventure. RSVP-TE extensions for interdomain LSPs. Technical Report 2002-09, University of Namur, October 2002. Available at http://www.infonet.fundp.ac.be/doc/tr/Infonet-TR-2002-09.html.

[13] Y. Rekhter and E. Rosen. Carrying label information in bgp-4, May 2001. RFC 3107.

[14] E. C. Rosen et al. BGP/MPLS VPNs, October 2002. Work in progress, draft-ietf-ppvpn-rfc2547bis-03.txt.

[15] V. Sharma and F. Hellstrand. Framework for MPLS-based recovery, October 2002. Work in progress, draft-ietf-mpls-recovery-frmwrk-08.txt.

[16] J.-P. Vasseur, C. Iturralde, R. Zhang, X. Vinet, S. Matsushima, and A. Atlas. RSVP Path computation request and reply messages, June 2002. Work in progress, draft-vasseur-mpls-computation-rsvp-03.txt.

[17] L. Xiao, K.-S. Lui, J. Wang, and K. Nahrstedt. QoS extension to BGP. In *10th IEEE International Conference on Network Protocols*, Paris, France, November 12-15 2002.

[18] R. Zhang et al. MPLS Inter-AS traffic engineering requirements, February 2003. Work in progress, draft-zhang-mpls-interas-te-req-02.txt.