Inter-Domain MPLS Restoration

Changcheng Huang, Donald Messier
Carleton University, 1125 Colonel By Drive, Ottawa, ON K1S 5B6 Canada
E-mail: huang@sce.carleton.ca

Abstract - **Several MPLS (Multi-Protocol Label Switching) based protection mechanisms have been proposed, such as end-to-end path protection and local repair mechanism. Those mechanisms are designed for intra-domain recoveries and little or no attention has been given to the case of crossing non-homogenous independent domains. This paper presents a novel solution for the setup and maintenance of independent protection mechanisms within individual domains and merged at the domain boundaries. This innovative solution offers significant advantages including fast recovery across multiple non-homogeneous domains and high scalability. Simulation results using OPNET are presented showing the advantage and feasibility of our proposed approach.**

Index Terms - **Protection, MPLS, Inter-Domain**

I. INTRODUCTION

The latency in Internet path failure, failover, and repair has been well documented over the years. This is especially true in the inter-domain case due to excessively long convergence properties of the Border Gateway Protocol (BGP) [1]. Research by C. Labovitz et al. [2] presents results, supported by a two year study, demonstrating the delay in Internet inter-domain path failovers averaged three minutes and some percentage of failover recoveries triggered routing table fluctuations lasting up to fifteen minutes. Furthermore the report states that "Internet path failover has significant deleterious impact on end-to-end performance-measured packet loss grows by a factor of 30 and latency by a factor of four during path restore". Although networks are becoming more and more resilient there are still frequent network failures that are becoming a cause for concern [3][4][5][6]. Future optical networks will carry a tremendous amount of traffic. A failure in an optical network will have a disastrous effect. The FCC has reported that network failures in the United States with impact on more than 30,000 customers happen with a frequency in the order of one every two days and the mean time to repair them is in the order of five to ten hours [7]. With optical communications technologies the problem is made worse and now a single failure may affect millions of users. Strategic planning at Gartner Group suggests at [8] that through 2004, large U.S. enterprises will have lost more than $500 million in potential revenue due to network failures that affect critical business functions. This Internet path failover latency is one of the driving factors behind advances in MPLS protection mechanism.

Protection and restoration issues have been widely studied under various contexts such as SONET rings, ATM and optical networks[9][10][11]. Several recovery mechanisms have been proposed over the last few years. End-to-end schemes provide protection on disjoint paths from source to destination and may rely on fault signaling to effect recovery switching at the source[12]. Local repair mechanisms for their part effect protection switching at the upstream node from the point of failure, the point of local repair (PLR) and do not require fault signaling[13][14]. Local repair has the advantage of fast recovery, but in general is not efficient in capacity. Path protection, on the other hand, can optimize spare capacity allocation on an end-to-end basis. Therefore it is typically more efficient.

MPLS being a relatively new technology, the research in advanced protection mechanism for MPLS is still in its infancy. This is especially true for inter-domain protection mechanism. The research conducted and still ongoing has identified several possible solutions to the MPLS intra-domain recovery problem[15]. Each of those solutions presents its own strengths and weaknesses. As a first cut, MPLS restoration schemes can be separated into on-demand and pre-established mechanism. On-demand mechanism relies on the establishments of new paths after the failure event while pre-established mechanism computes and maintains restoration paths for the duration of the communication session. Due to the fast recovery times sought, this work focuses exclusively on pre-established protection switching. Of the pre-established recovery mechanisms, one of the first being implemented in a commercial product is Cisco Systems' Fast Re-route (FRR) algorithm in the Gigabit Switch Router family. FRR provides very fast link failure protection and is based on the establishment of pre-established bypass tunnels for all Label Switch Routers. The FRR algorithm can switch traffic on a failed link to a recovery path within 20 ms but is limited to the global label assignment case. Several other methods have been proposed based on individual backup LSPs (Label Switched Path) established on disjoint paths from source to destination. An immediate benefit of end-to-end mechanism is scalability. Reference [16] shows that given a network of N nodes, local repair schemes require

*N*L*(L-1)* backup paths to protect a network if each node has *L* bi-directional links. For end-to-end schemes, a network with *M* edge nodes, the total number of backup paths is proportional to *M*(M-1)*. If M is kept small, a significant scalability advantage is realized. The following paragraphs provide an overview of the most promising intra-domain protection schemes.

The proposal at [17] is an improvement over the one hop CISCO FRR and describes mechanisms to locally recover from link and node failures. Several extensions to RSVP-TE are introduced to enable appropriate signaling for the establishment, maintenance and switchover operations of bypass tunnels and detour paths. The Fast Reroute method will be referred to as Local Fast Reroute (LFR) in this paper. In the Local Fast Reroute, one-to-one backup LSPs can be established to locally bypass a point of failure.

A key part of this proposal is to setup backup LSPs by making use of a label stack. Instead of creating a separate LSP for every backed-up LSP, a single LSP is created which serves to backup a set of LSPs. Such an LSP backing up a set of primary LSPs is called a bypass tunnel.

The key advantage of LFR is the very fast recovery time while its disadvantages are scalability issues due to the potential large number of bi-directional links and complexity in maintaining all the necessary label associations for the various protected paths.

The first end-to-end path protection scheme is presented at [16] and uses signaling from the point of failure to inform the upstream LSRs (Label Switching Router) that a path has failed. Here a Reverse Notification Tree (RNT) is established and maintained to distribute the fault and recovery notifications to all ingress nodes which may be hidden due to label merging operations along the path. The RNT is based on the establishment of a Path Switch LSR (PSL) and a Path Merge LSR (PML). The PSL is the origin of the recovery path while the PML is its destination. In the case of multipoint-to-point tree the PSLs form leaves and the PMLs roots of multicast trees. The main advantages of RNT protection are scalability and efficient use of resources while its disadvantages are long recovery time due to the propagation of failure notification messages.

Another end-to-end path protection mechanism presented at [18] is called End-to-end Fast Reroute (EFR). It can achieve nearly the same protection speed as LFR, but is extremely inefficient in terms of bandwidth resource. It requires about two times of the bandwidth of the protected path for protection path. For more about this approach, readers are referred to [18].

Current research and recent proposals deal with the intra-domain case or assume homogeneity and full cooperation between domains. Recognizing the growing need to provide a solution for the more general case, this paper proposes a new and innovative solution to solve the inter-domain protection problem for LSPs spanning multiple inhomogeneous and independent domains. The proposed solution is based on the use of concatenated primary and backup LSPs, protection signaling and a domain boundary protection scheme using local repair bypass tunnels. We will call our scheme Inter-Domain Boundary Local Bypass Tunnel (IBLBT) in this paper to differentiate it with other solutions.

II. PROBLEM STATEMENT AND PROPOSED SOLUTION

The MPLS protection mechanisms presented in Section I include LFR, EFR and RNT. All were designed for intra-domain failure recovery and will generally not function when the primary LSP is not bounded to a single administrative domain. The scalability problem with LFR will be stretched further if multiple domains are involved because each domain may have hundreds of nodes and links that require bypass tunnels for protection. While both EFR and RNT suffer longer delays due to the long LSPs that span several domains, EFR becomes more inefficient compared to RNT because its extra bandwidth requirements.

A unique issue for inter-domain protection is that, separate domains may not cooperate with each other. Each domain is administered through a different authority. Some authorities, such as carriers, are not willing to share information with each other. Certain critical information may have significant impact on the operation of public carriers if it is disclosed. For example, failure information is typically considered negative on the image of a public carrier and can be exploited by competitors for their advantages. Most carriers will consider this information confidential and will not likely share this information with their customers and other carriers. When an internal failure happens, a carrier will try to contain this information within its own domain and try to recover from the failure by itself. Both end-to-end RNT and end-to-end EFR require some kind of failure signaling to all the upstream domains. Containing this failure signaling to the originating domain will make end-to-end RNT and EFR almost impossible.

A complete solution to the inter-domain protection problem can be found if we turn the apparent difficulties in end-to-end RNT into advantages. Such is the case for the independence of domains. Accepting the fact that domains will be independent and inhomogeneous leads to the idea of establishing an independent path protection mechanism within each domain while at the same time being able to guarantee protection throughout the path from end to end. What is required is a solution at the domain boundaries to ensure protection continuity. For the solution to work, each domain must provide its own RNT protection scheme which it initiates, establishes, maintains and hands over to the next protection domain at the domain boundary. A domain protection scheme must therefore be executed completely within that domain with no involvement from other domains. The first step towards this solution is to

allow the primary and backup LSPs to be concatenated at the domain boundaries. Concatenation in this context is used to mean that specific actions must be taken at this point in the LSP to ensure continuity of service and protection across domain boundaries. Figure 1 illustrates the fundamental principles behind this solution. The primary path P1 is protected through three separate backup paths namely B1, B2 and B3. B1 is initiated in the source domain, B2 at the domain boundary and B3 in the destination domain. Each of those backup paths is independent of the others and does not require fault notification beyond its own domain.

This innovative solution permits the isolation of the protection mechanism to a single domain or domain boundary. Furthermore, domains can now use independent protection mechanisms and signaling schemes and do not need to propagate their internal failure notifications to adjacent domains. This solution combines the advantages of fast local repair at the domain boundaries and the scalability advantages of end-to-end protection within domains.

In summary, the proposed solution to the inter-domain MPLS recovery problem is based on the establishment of independent protection mechanisms within domains using concatenated primary and backup LSPs, minimal protection signaling between domains and, local repair at the domain boundaries. Viewed from end-to-end at figure 1, the primary LSP is protected by three or more distinct and independent protection regions merged at their respective boundaries. Those protection regions are the Source Protection Domain, the Domain Interface Protection and the Destination/Transit Protection Domain. In addition to those three protection regions, transit protection regions are also possible when a protected LSP transits one or more independent domains before reaching its destination. In such a case, there would be several domain interface protections in place.

Our solution introduces and makes use of Gateway LSRs and Concatenation Path Switch LSRs (CPSLs) as well as Proxy Concatenation PSLs (PCPSL) and Proxy Gateway LSRs (PGL). Those new protection elements are used to pre-establish inter-domain local bypass tunnels and guarantee protection against node and link failures when sufficient protection elements are present.

In the following discussions, we assume that there are at least two separate links connecting two pairs of border routers between any neighboring domains. This will allow us to provide protection for two neighboring domains without counting on the support of a third domain under the context of single point failure. One example that satisfies this requirement is shown in figure 2. Our focus is therefore on removing the interdependency among domains that are not directly linked and further limiting the dependency between neighboring domains as discussed in the next section. We will use the scenario of Dual Exit LSRs Fully Meshed (figure 2) as our example case. The principles of our solution can be readily applied to all other scenarios that satisfy the condition stated at the beginning of this paragraph.

Figure 2 illustrates the topology where a primary protected LSP P1A is protected in Domain A via backup path B1A, protected at the boundary via local backup path B1 and protected in Domain B through backup path B1B. LSR 0 is the selected Gateway LSR for path P1 while LSR 1 is its corresponding PGL. LSR 2 is the CPSL for the same primary path while LSR 3 is the PCPSL. The PGL and PCPSL are responsible to maintain end-to-end path integrity in the event of a Gateway or CPSL failure. The selection of the PCPSL and its significance in the recovery process is critical for the operation of this scheme. This point is evident when looking at figure 2. In the figure we note that B1 and B1B are routed through the PCPSL LSR 3. Although the identification of the PCPSL is simple in a Dual Exit LSR topology its role is nevertheless important. It is the merging of the local inter-domain backup path B1 and the destination domain backup path B1B at the PCPSL LSR 3 that permits full and continuous protection across domains. Without this action, recovery traffic on B1 would be dropped at LSR 3 since it could not make the necessary label association. The merging action of the PCPSL ensures label binding between B1 and B1B, enabling the recovery traffic from the Gateway LSR to be switched to the destination.

III. SIMULATION RESULTS

To verify the potential for the proposed IBLBT solution, two separate models were built. The first model implements MPLS recovery using an end-to-end path protection mechanism. The model was built using dynamic LSPs. For the end-to-end recovery to work it is necessary for all nodes in the model to share a common signaling and recovery mechanism. This is necessary in the extended intra-domain end-to-end scheme since domains have to fully cooperate in the recovery process. As discussed in previous chapters, this naïve extended intra-domain solution would likely not be found in real networks. Nevertheless, the model is useful to serve as a comparison point with IBLBT solution proposed in this paper. In contrast to end-to-end recovery, IBLBT isolates recovery to the domain boundary or to an individual domain. The second model built is the model implementing the proposed solution with its inter-domain boundary local repair tunnels. All models are run with various failure scenarios to collect data on recovery time for further analysis and comparison.

A. MPLS End-To-End Protection Model

The first MPLS model was built to measure recovery time for an end-to-end protection case and is represented at figure 3. As stated earlier, it is recognized that such an inter-domain end-to-end protection mechanism is naïve for the reasons discussed in Section II. However, to obtain comparative data from such a scheme LSPs were configured using dynamic LSPs and all nodes share the

same signaling and protection mechanism. Traffic was generated using two separate Gigabit Ethernet LANs each with twenty-five users running high-resolution video conferencing applications over UDP. Additional applications were configured such as heavy database access and email, file transfers, and print sessions using TCP. Traffic entered the MPLS network at the ingress node LSR 0. The Egress and Ingress LSRs were modeled as CISCO 7609 while the transit LSRs were CISCO 7000 routers. The Egress and Ingress LSRs were selected based on the number of Gigabit Ethernet ports available for the source LANs. IP forwarding processor speeds were increased to 50000 packets/sec on all nodes to permit higher traffic volumes for the simulation. High traffic volume was necessary to ensure high link utilization for measurement purposes. Traffic was switched between LSRs based on the Forward Equivalency Class (FEC) associated with the incoming traffic and the established Paths. The selected Primary Path is shown at figure 3 and follows path LSRs 0-1-4-5-7-8 while the pre-established end-to-end backup LSP follows LSRs 1-3-6-7-8 (LSR 1 is the PSL). All model links are OC-12 with 0.8 ms delay for inter-domain links and 4 ms delay for intra-domain links. This approximates 1200 km intra-domain links and 240 km inter-domain links. The average load on the network was kept at approximately 125 Mbps.

Several failure scenarios were studied as follows:

a) Source domain failure (Link 1-4 failure);
b) Domain interface Failure (Link 4-5 and node 5 failure);
c) Destination domain failure (Link 5-7 failure).

A failure and recovery process was configured in Opnet to effect the planned failures at 170 sec from the simulation start time. All simulations were run for a total of 180 seconds. The 170 seconds time before failure was selected to ensure sufficient time for all routing processes to complete their initial convergence, for traffic generation processes to reach steady state prior to the network failure, and for the MPLS processes to establish LSPs after the initial layer three routing protocol convergence. The simulation end time is selected to allow sufficient time for recovery and steady states to return while being kept at a minimum to reduce the simulation run time. The large amount of application traffic generated during the simulation caused the simulation run time to be in excess of one hour.

This model makes use of CR-LDP keep-alive messages to detect node failures while link failures are detected through lower layer Loss of Signal (LOS). The keep-alive message interval was configured for 10 ms while the hold off timer was set at 30 ms. Those short intervals were selected arbitrarily but taking into account the OC-12 line rate with a view to reduce packet loss during failover. Upon detecting a failure the node upstream from the point of failure sends an LDP notification message to the source node informing it of the failure and the affected LSPs. Base on this notification message, the source node switches to the recovery path. This LDP notification message is encapsulated in an IP packet and forwarded to the ingress node for action. Several network probes were configured to collect data on recovery times, routing tables, link state databases, traffic in and out of all LSPs as well as forwarding buffer utilization.

For this work, recovery time was measured at the merge point of the primary and backup paths (ie: PML). This recovery time is from the receiver's perspective and represents the difference in time between the reception of the last packets on the primary path and reception of the first packets on the recovery path. The recovery time includes failure detection time, time for the transmission of failure notification messages, protection-switching time, and transmission delay from the recovery point to the merge point. To obtain the necessary LSP traffic data for the measurement of recovery time, LSP output traffic for primary and backup LSPs at the merge point was sample every 100 μsec. This sampling time was selected to provide sufficient granularity into the recovery process while maintaining simulation output files to a manageable size.

B. Inter-Domain Boundary Bypass Tunnel Model

In the IBLBT model, the primary and backup paths were established following the proposed inter-domain protection algorithms. As described in previous sections and depicted at figure 4, concatenated LSPs were setup within each domain with backup paths using bypass tunnels established manually as described in section II. The simulations were run for 130 seconds with failures programmed for 125 seconds. Shorter simulation time is possible with this model because static LSPs are used and no setup time is required during the simulation. Other than the recovery mechanism, the model was setup identically to the previous end-to-end MPLS model.

C. Results Summary

Comparing end-to-end recovery with the IBLBT case is shown in table 1 (results are different with statistically significance). The recovery speed benefits of IBLBT over the end-to-end case would have been much more evident had the simulation model included several more independent domains. Of course the further away the failure is from the point of repair the longer the recovery time. Given the simplicity of the models in this work, the significant advantages of IBLBT could not be exploited fully against the end-to-end case.

IV. CONCLUSIONS

The growing demand for QoS has led to significant innovations and improvements on the traditional best effort IP networks. Technologies such as MPLS provide important advantages over the classical hop-by-hop routing decision processes. The ability of MPLS to apply equally well to various layer 1 technologies including Wavelength

Division Multiplexing (WDM) makes this technology a strong contender for current leading edge and future networks. Furthermore, due to its label switching architecture, MPLS can provide very fast recovery mechanism complementing existing lower layer protection schemes. The development of new techniques to provide path protection at the MPLS layer will certainly continue. The proposed IBLBT protection mechanism presented in this paper is an innovative and unique scheme to provide protection across multiple independent domains. It relies on only a very basic amount of information provided by neighboring domains and makes no assumption on protection mechanisms of other domains and level of cooperation. Simulation results show recovery times of a few milliseconds and point to the potential for this proposed solution for MPLS inter-domain protection.

In general, our solution permits the isolation of the protection mechanism to a single domain or domain boundary. Furthermore, domains can now use independent protection mechanisms and signaling schemes and do not need to propagate their internal failure notifications to adjacent domains. This solution combines the advantages of fast local repair at the domain boundaries and the scalability advantages of path protection within domains.

REFERENCES

[1] Y. Rekhter, T.J. Watson, T. Li, "A Border Gateway Protocol 4 (BGP-4)," *IETF RFC 1771*, March 1995.

[2] C. Labovits, A. Ahuja, A. Bose, F. Jahanian "Delayed Internet Routing Convergence," *IEEE/ACM Transactions on Networking*, Vol. 9, No. 3, June 2001.

[3] T. G. Griffin, G. Wilfong, " An Analysis of BGP Convergence Properties," in *ACM SIGCOMM 1999*, Cambridge, September 1999.

[4] S. J. Jeong, C. H. Youn, T. S. Choi, T.S. Jeong, D. Lee, K.S. Min, " Policy Management for BGP Routing Convergence Using Inter-AS Relationship," *Journal of Communications and Networks*, Vol. 3, No.4, December 2001.

[5] R. Mahajan, D. Wetherall, and T. Anderson, "Understanding BGP Misconfiguration," in *ACM SIGCOMM 2002*, Pittsburgh, August 2000.

[6] K. Varadhan, R. Govindan, and D. Estrin, "Persistent Route Oscillations in Inter-Domain Routing," *Computer Networks*, 32(1), 1999.

[7] P. Demeester, T. Wu, N. Yoshikai, "Survivable Communications Networks," *IEEE Communications*, Vol. 37 No. 8, August 1999.

[8] B. Hafner, J. Pultz, "Network Failures: Be afraid, be very afraid," *Gartner Group, Note SPA-09-1285*, 15 September 1999.

[9] O. J. Wasem, "An Algorithm for Designing Rings for Survivable Fiber Networks," *IEEE Transactions on Reliability*, vol. 40, 1991.

[10] T. Frisanco, "Optimal Spare Capacity Design for Various Switching Methods in ATM Networks," *ICC'97*, Montreal, June 1997.

[11] B. T. Doshi, S. Dravida, P. Harshavardhana, O. Hauser, and Y. Wang, "Optical Network Design and Restoration," *Bell Labs Technical Journal*, January-March 1999.

[12] P.-H. Ho and H. T. Mouftah, "Reconfiguration of Spare Capacity for MPLS-based Recovery," *To appear in IEEE/ACM Transaction on Networking.*

[13] M. Kodialam and T. V. Laksman, "Dynamic Routing of Locally Restorable Bandwidth Guaranteed Tunnels Using Aggregate Link Information," *IEEE Infocom'01*, Anchorage, April, 2001.

[14] W. D. Grover and D. Stamatelakis, "Cycle-Oriented Distributed Preconfiguration: Ring-like Speed with Mesh-like Capacity for Self-planning Network Restoration," *IEEE ICC'98*, Dresden, June 1998

[15] V. Sharma and F. Hellstrand, "Framework for Multi-Protocol Label Switching (MPLS) Based Recovery," *IETF RFC 3469*, February, 2003.

[16] C. Huang, V. Sharma, S. Makam, K. Owens, "Building Reliable MPLS Networks Using a Path Protection Mechanism," *IEEE Communications Magazine*, March, 2002.

[17] P. Pan, D.H. Gan, G. Swallow, J.P. Vasseur, D. Cooper, A. Atlas, M. Jork, "Fast Reroute Extensions to RSVP-TE for LSP Tunnels", IETF Draft, Work in progress, <draft-ietf-mpls-rsvp-lsp-fastreroute-00.txt>, Jan 2002.

[18] D. Haskin and R. Krishnan, "A method for setting an alternative label switched paths to handle fast reroute", IETF Draft, Work in progress <draft-haskin-mpls-fast-reroute-05.txt>, November 2000.

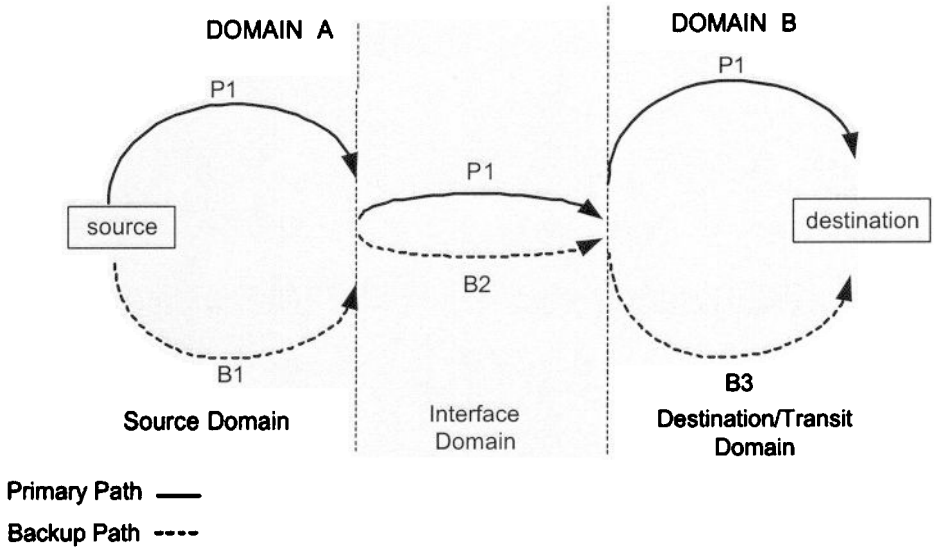

Figure 1 - Concatenated Primary and Backup LSPs

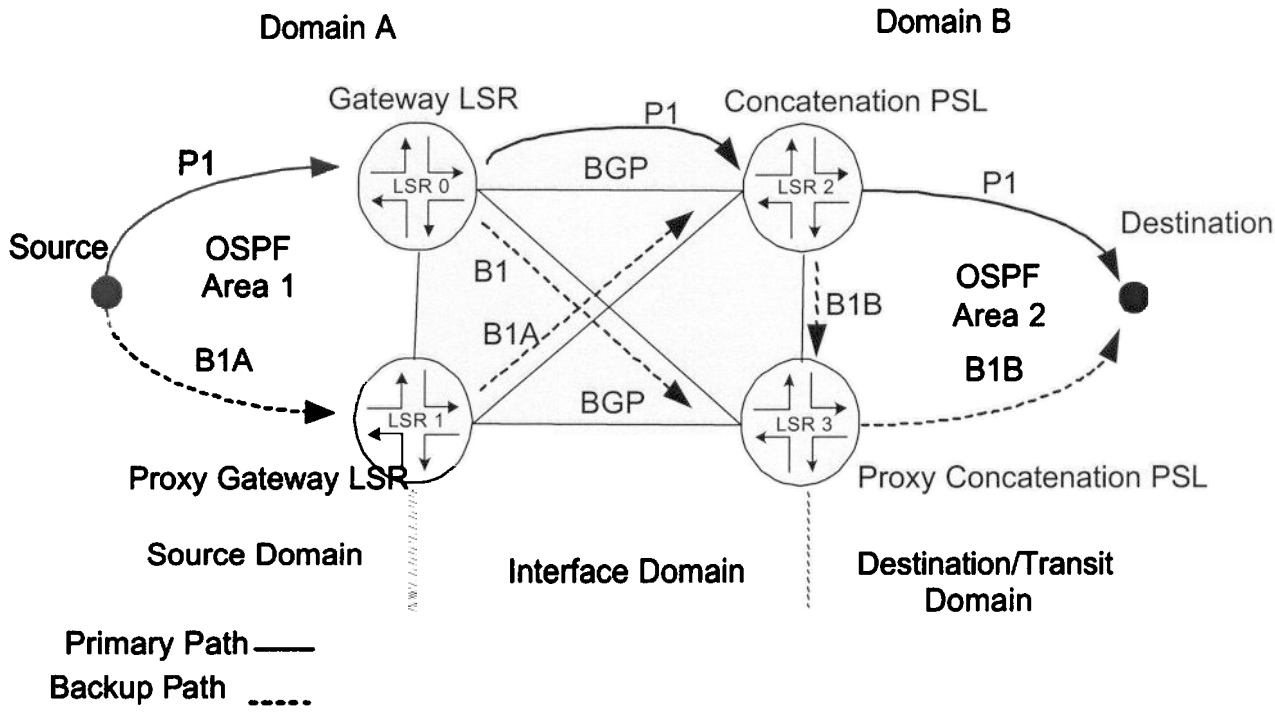

Figure 2 – Inter-Domain Protection

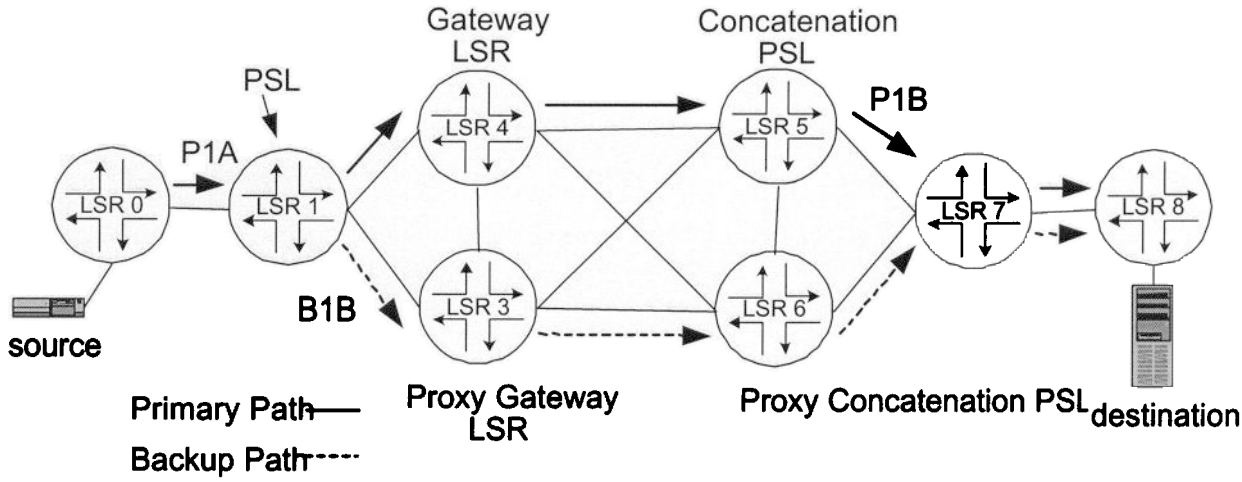

Figure 3 – MPLS End-to-end Protection Model

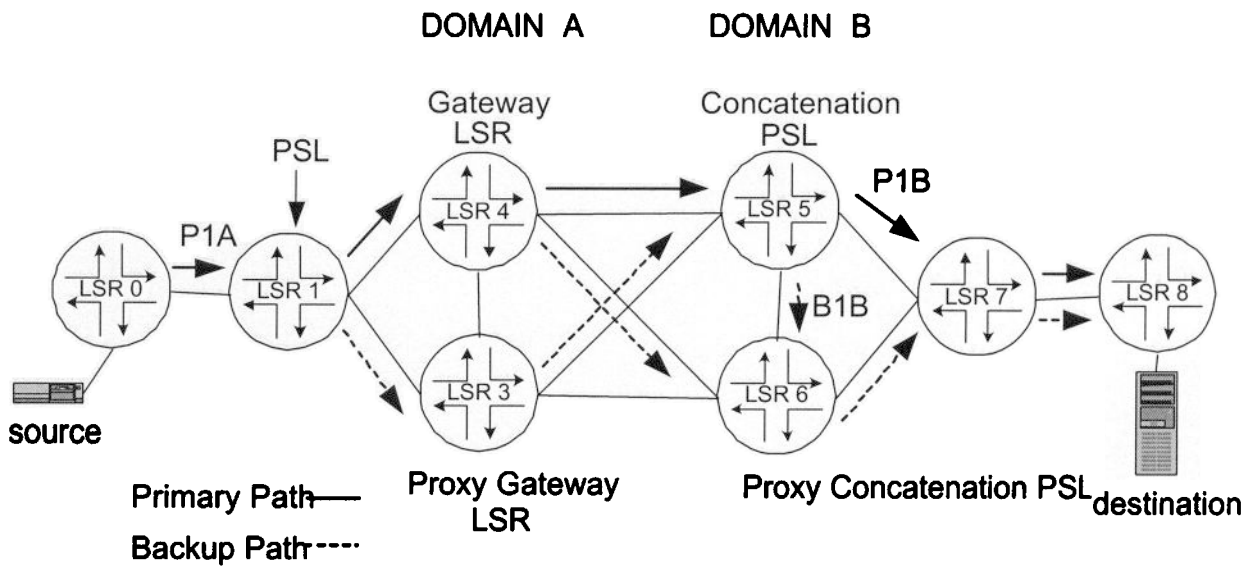

Figure 4 – MPLS Domain Boundary Local Bypass Tunnel Model

TABLE I – COMPARISONS OF END-TO-END RECOVERY AND IBLBT

	Link 1-4	Link 4-5	Link 5-7	Node 5
IBLBT	4.7 ms	1.19 ms	8.56 ms	32.02 ms
End-to-end recovery	7.18 ms	12.0 ms	13.24 ms	46.38 ms