**MPLS Inter-Domain Protection
using Domain Boundary Local Bypass Tunnels**

A Thesis submitted to the faculty of Graduate Studies and Research in partial fulfillment
of the requirements for the degree of

Masters of Engineering

Ottawa Carleton Institute for Electrical and Computer Engineering

Department of Systems and Computer Engineering

Carleton University

Ottawa, Canada

July 2002

Submitted by

Donald Messier

Student Nr: 257007

BEng, Royal Military College, Kingston, Ontario, Canada

National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-79771-6

Canada

The undersigned recommends to the Faculty of Graduate Studies and Research
acceptance of the Thesis

**“MPLS Inter-Domain Protection
using Domain Boundary Local Bypass Tunnels”**

Submitted by

Donald Messier

257007

BEng, Royal Military College of Canada, Kingston, Ontario

In partial fulfillment of the requirements for the degree of Masters of Engineering.

C. Huang
Thesis Supervisor

Chair
Department of Systems and Computer Engineering
Carleton University
Ottawa, Canada

To my friend Rita

This document is the culmination of two years of studies and research. During that time we have been through a lot together. It has been hard, it has been good, it has been fun, it has been painful, it most certainly has been enlightening. I only wish you were still physically here to celebrate with me. As you know I have been very busy, but I promise, I will see you and talk to you soon.

ABSTRACT

Multi-Protocol Label Switching (MPLS) technology is based on the separation of routing and forwarding processes. MPLS can provide much faster forwarding by switching traffic using labels. Efficiencies are achieved at intermediate nodes by avoiding hop by hop layer three address resolution. Furthermore, MPLS enables criteria based traffic engineering for various traffic granularity. The widespread interest and unparalleled potential of MPLS is leading to further research and development efforts. One of those area of research is Path Protection Mechanism. It is widely accepted that layer three protection and recovery mechanism are too slow for today's QoS requirements. Failure recovery latencies of several seconds to minutes for layer three routing protocols have been widely reported. For this reason, a recovery mechanism at the MPLS layer capable of recovering from failed paths in 10's of milliseconds has been sought. In this light, several MPLS based protection mechanism have been proposed and include End-to-end Fast Reroute, End-to-end recovery with Reverse Notification Tree and, Local Repair mechanism. Those mechanism are designed for intra-domain recoveries and little or no attention has been given to the case of non-homogenous independent inter-domains. This thesis presents a solution for the setup and maintenance of independent protection mechanism within individual domains and merged at the domain boundaries. This innovative solution offers significant advantages including fast recovery across multiple non-homogeneous domains. Detail setup and operations procedures are described as well as protocol and process modifications. Finally, simulation results using Opnet are presented showing recovery times of a few milliseconds.

TABLE OF CONTENT

		Page
1	Chapter 1 Introduction	7
1.1	Background	7
1.2	MPLS Overview	9
1.3	BGP Overview	12
1.4	Research Summary	14
1.5	Thesis Aim and Contributions	15
1.6	Limitations	15
1.7	Thesis Organization	16
2	Chapter 2 State of the Art	17
2.2	Local Fast Reroute	19
2.3	End-to-end Fast Reroute	21
2.4	Reverse Notification Tree	23
3	Chapter 3 Problem Statement	25
3.2	Single Exit LSR	28
3.3	Single Exit LSR Multi-homed	29
3.4	Single Exit LSR Multi-homed, Multi-domain	30
3.5	Dual Exit LSR	31
3.6	Dual Fully Meshed Exit LSR	32
3.7	Summary	33
4	Chapter 4 Proposed Solution	35
4.2	Inter-Domain Boundary Local Bypass Tunnel	39

4.3	Source Domain Protection Setup	43
4.4	Destination/Transit Domain Protection Setup	46
4.5	Summary.....	47
4.6	Fault Detection	49
5	Chapter 5 Path Recovery Procedures.....	51
5.2	Gateway LSR Procedures	51
5.3	Proxy Gateway Procedures	54
5.4	Concatenation PSL Procedures	55
5.5	Proxy Concatenation PSL Procedures.....	56
6	Chapter 6 Bandwidth Reservation Issues	57
6.1	General	57
7	Chapter 7 Advantages	61
7.1	General	61
8	Chapter 8 Simulation and Results	63
8.2	BGP Model	64
8.3	MPLS End-to-End Recovery Model	68
8.5	MPLS Inter-Domain Boundary Local Repair Model	75
8.7	Results Summary	80
9	Chapter 9 Conclusion... ..	81
9.2	Advantages	82
9.3	Disadvantages	83
9.4	Further Research	83

Terminology	85
List of Acronyms	89
References	90

Appendix A – Implementation with RSVP-TE

Appendix B – Modifications and Extensions

LIST OF FIGURES

		Page
Figure 1	Label Switching	11
Figure 2	The Border Gateway Protocol	13
Figure 3	Local Fast Reroute	19
Figure 4	End-to-end Fast Reroute	22
Figure 5	End-to-end Protection with RNT	23
Figure 6	The Inter-Domain Protection Problem	25
Figure 7a	Single Exit LSR	29
Figure 7b	Single Exit LSR, Multi-homed	30
Figure 7c	Single Exit LSR, Multi-homed, Multi-domain	31
Figure 7d	Dual Exit LSR	32
Figure 7e	Dual Exit LSR, Fully Meshed	33
Figure 8	Concatenated Primary and Backup LSPs	36
Figure 9	Inter-Domain Protection	38
Figure 10	Fully Meshed Dual Exit LSRs	41
Figure 11	Dual Exit LSRs	42
Figure 12	Source Domain Protection Setup for Fully Meshed Dual Exit LSRs	44
Figure 13	Source Domain Protection Setup for Dual Exit LSRs	45
Figure 14	Destination/Transit Domain Protection Setup	46
Figure 15	Pre-failure steady state	48
Figure 16	Pre-failure steady state for Dual Exit LSRs	49
Figure 17	Inter-Domain Link or Concatenation PSL failure Recovery ..	52
Figure 18	Source Domain Failure Recovery	54
Figure 19	Destination/Transit Domain Failure Recovery	55
Figure 20	Layer 3 BGP 4 Baseline Model	65
Figure 21	Received IP packets, BGP 4 model	67
Figure 22	End-to-end MPLS Protection Model	69
Figure 23	Received IP Packets, End-to-end MPLS model	74

Figure 24	Primary Path output, end-to-end MPLS recovery model	74
Figure 25	Backup Path output, end-to-end MPLS recovery model	75
Figure 26	MPLS Domain Boundary Local Bypass Tunnel Model	76
Figure 27	Primary Path output, local repair MPLS model	78
Figure 28	Backup Path output, local repair MPLS model	78

CHAPTER 1 - INTRODUCTION

1.1 BACKGROUND

1.1.1 Real-time communication services have become increasingly critical to businesses and governments. Unlike traditional datagram services where quality of service (QoS) is measured in terms of availability and throughput, real time services require additional QoS criteria such as delay, delay variations, cell drop rate, error rate and increasingly, fault tolerance and fast recovery. Those QoS requirements combined with the convergence of networking technologies towards an IP based infrastructure has placed significant new demands on the traditional best effort IP network. As a result, several new technologies have emerged to improve the QoS of the IP domain and enable real time applications to converge to this domain. Multi Protocol Label Switching (MPLS) is one such technology enabling improved QoS control, granularity and traffic engineering in an IP or other network layer domain. Two signaling protocols are defined to bind MPLS labels to Forward Equivalency Class (FEC) and distribute those labels to MPLS peers. First, the Label Distribution Protocol (LDP) designed specifically for this task is described in RFC 3036. Second, the Resource Reservation Protocol (RSVP) defined in RFC 2205 and extended to support label binding and distribution with RSVP-TE [4]. MPLS, when combined with a differentiated services model, offers faster switching operations and traffic engineering with pre-determined and guaranteed QoS using pre-established and reserved Label Switch Paths (LSP). This improved QoS offering opens the IP door to potential customers traditionally bound, due to their high QoS demands, to circuit switch and connection oriented services. Unfortunately, the slow

fault recovery mechanism of IP which can take several seconds to minutes to recover from a failure is still keeping the IP door closed for some service providers, applications and users who can not tolerate such long outages. As a result, a mechanism to quickly recover from failures at the MPLS layer has been sought to complement existing lower layer recovery mechanism. The goal is to provide a recovery mechanism at the MPLS layer capable of restoring services around a failure point in tens of milliseconds (10's ms). This fast recovery time would be comparable to SONET recovery as specified in GR253.

1.1.2 Several MPLS recovery mechanism 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 or at a pre-determined Path Switch Label Switch Router (PSL). 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. 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 non homogeneous 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.

1.2 MPLS OVERVIEW

1.2.1 As a packet of connectionless network layer protocol travels from one router to the next, each router makes an independent forwarding decision for that packet. That is, each router runs a network layer routing algorithm. Each router independently chooses a next hop for the packet, based on its analysis of the packet's header and the results of running the routing algorithm. Packet headers contain considerably more information than is needed simply to choose the next hop. Choosing the next hop can therefore be thought of as the composition of two functions. The first function partitions the entire set of possible packets into a set of "Forwarding Equivalence Classes (FECs)". The second maps each FEC to a next hop. Insofar as the forwarding decision is concerned, different packets which get mapped into the same FEC are indistinguishable. All packets which belong to a particular FEC and which travel from a particular node will follow the same path, or if certain kinds of multi-path routing are in use, they will all follow one of a set of paths associated with the FEC. [24]

1.2.2 In MPLS, the assignment of a particular packet to a particular FEC is done just once, as the packet enters the MPLS network at the ingress LSR. That FEC can be assigned based on many criteria such as destination address, QoS requirement, incoming interface, incoming port, etc. The FEC assigned to the packet is encoded into a short fixed label on top of the packet. The packet is then forwarded to the next hop. At subsequent hops, there is no further examination of the layer three addresses. The MPLS

label is used as an index into a table which contains the next hop and new label for the label swap operation.

1.2.3 Figure 1 depicts a typical IP domain with MPLS capable routers. Source 1 and source 2 packets enter the MPLS network at ingress LSRs 3 and 1 respectively. Sometimes those ingress Label Switch Routers (LSRs) are referred to as Label Edge Routers (LERs). Source 1 generates traffic for destination 4 and destination 2 while source 2 generates traffic for destination 4 only. The ingress LSRs assign FECs to the incoming traffic. Each FEC is associated to a label. In figure 1, LSR 1 assigns label 56 to traffic for destination 4 while LSR 3 assigns label 5 for the same traffic. Label assignments are distributed in the network using a signaling protocol such as RSVP-TE or CR-LDP. The next hop along the path, LSR4 for example, simply examines the incoming label, looks up the index, swaps the incoming label with the outgoing label 35 and switches the packet towards LSR 5. LSR 5 recognizing that traffic from source 2 and source 1 have the same FEC, merges the two for destination 4 into a single outgoing label, label 23. The label merge operation is performed during the initial path setup. LSR 16, the egress LSR, removes (Pops) the labels and forwards the packet to destination 4. LSR 15 may also pop the label and forward the packet to LSR 16 unlabelled since it is the penultimate hop for that path.

Figure 1 Label Switching

1.2.4 MPLS label stack permits the formation of path hierarchies. The label stack is the ordering of labels into a last in first out scheme. In figure 1 for example, a bypass tunnel from LSR 5 to LSR 15 is established using a stacked label. The reason for the bypass tunnel may be to provide protection for the primary path or provide a traffic engineered alternate path when congestion occurs on the primary path. In those cases, LSR 5 removes the incoming label and pushes label 76 into the label stack. Assuming global label space, Label 76 is the expected label at LSR 15, the tunnel end point. LSR 5 then pushes the tunnel label, label 25, onto the stack. The label stack now contains labels 76 and 25. The packets follows the bypass tunnel path and reaches LSR 5 which proceeds to remove the tunnel label and expose label 76. The label stack technique can also be used

to distribute label attributes. For example, given that egress LSR 13 supports three destinations, it would normally have to pop the last label and examine the layer 3 addresses of the incoming packets prior to determine the appropriate destination. LSR 13 could however, distribute a label attribute for each of those destinations. In figure 1, a label attribute of 2 is distributed to refer to destination 2. In this case, the ingress LSR 3, pushes attribute label 2 onto the stack and then pushes the primary path label, label 12. In this way, LSR 13 no longer needs to examine the layer three destination address before forwarding the packets to destination 2.

1.3 BORDER GATEWAY PROTOCOL OVERVIEW

1.3.1 The border Gateway Protocol (BGP) is an inter-domain routing protocol. It is built on experience gained with the Exterior Gateway Protocol (EGP). The primary function of BGP is to exchange network reachability information with other BGP speaking routers. In this way, independent Autonomous Systems (ASs) can exchange routing information without the need to share a common internal routing protocol. Figure 2 illustrates the BGP concept and depicts three ASs using different internal routing protocols such as IS-IS, RIP and OSPF. The ASs are connected and exchange reachability information using BGP. The Network reachability information includes information on the list of ASs that reachability information traverses. This information is sufficient to construct a graph of AS connectivity from which routing loops may be pruned and some policy decisions at the AS level may be enforced. BGP-4 is the latest version of BGP and provides a new set of mechanism for supporting classless inter-

domain routing. A BGP peer is a BGP speaking host capable of exchanging BGP routing information. In figure 2, Router 1 has two BGP peers, Routers 2 and 5. Similarly, router 5 has three peers, routers 1, 9 and 7. Routers 7 and 9 are internal BGP peers. [24]

1.3.2 BGP uses TCP for reliable transport protocol and operates by establishing a transport protocol connection between BGP peers. The peers exchange messages to open and confirm the connection parameters. The initial data flow is the entire BGP routing

Figure 2 The Border Gateway Protocol

table. Incremental updates are sent as the routing tables change. BGP does not require periodic refresh of the entire routing table. Therefore, a BGP speaker must retain the current version of the entire BGP routing tables of all its peers for the duration of the

connection. KeepAlive messages are sent periodically to ensure the liveliness of the connection. Notification messages are sent in response to errors or special conditions. If a particular AS has multiple BGP speakers and is providing transit service for other ASs, then care must be taken to ensure a consistent view of routing within the AS. A consistent view of the interior routes of the AS is provided by the interior routing protocol. A consistent view of the routes exterior to the AS can be provided by having all BGP speakers within the AS maintain direct BGP connections with each other such as depicted in figure 2. Using a common set of policies, the BGP speakers arrive at an agreement as to which border routers will serve as the exit/entry points for particular destinations outside the AS. This information is communicated to the AS's internal routers via the Interior Routing Protocol (IRP). Connections between BGP speakers of different ASs are referred to as "external" links while connections to BGP peers in the same AS are "internal" links. [24]

1.4 RESEARCH SUMMARY

1.4.1 In researching MPLS protection mechanism several sources have been surveyed including IEEE and IETF publications. Much of the research in this area deals with path optimization, resource sharing, protection path setup mechanism and signaling protocol modifications. Research on MPLS protection across multiple independent domains is new and no work was found related to this specific topic.

1.5 THESIS AIM AND CONTRIBUTION

1.5.1 The aim of this thesis is to propose a solution to solve the MPLS inter-domain protection problem and describe the necessary algorithms and protocol modifications to implement the proposed solution.

1.5.2 The contribution of this work is the introduction of an innovative inter-domain protection mechanism providing single link or node failure protection. The proposed mechanism permits the establishment and maintenance of protected LSPs and their associated backup LSPs using independent protection schemes within individual domains and a new domain boundary protection using local bypass tunnels. The scheme introduces new protection element functionality at the Gateway LSR, Proxy Gateway LSR (PGL), Concatenation PSL (CPSL) and Proxy Concatenation PSL (PCPSL). Several new algorithms are proposed along with extensions to existing protocols. The proposed inter-domain protection mechanism offers several important advantages such as the isolation of fault recovery within individual domains, fast local protection switching at the domain boundaries, protection mechanism independence and protection signaling independence.

1.6 LIMITATIONS

1.6.1 The case where single exit and entry LSRs exists between domains but multiple paths do exists through transit domains as depicted at figure 6, will not be addressed.

1.6.2 Border Gateway LSRs must share a common signaling protocol to exchange label information and QoS requirements.

1.6.3 The proposed protection mechanism is effective for a single path failure at a time.

1.6.4 Every domain along a path must provide a protection mechanism.

1.7 THESIS ORGANIZATION

1.7.1 This thesis is organized progressively starting with an overview of the current state of MPLS protection technologies. This discussion is followed by an analysis of the problem and scoping of the problem. The details of the proposed inter-domain protection solution are then presented including. Various failure scenarios are analyzed with the proposed protection solution implementation. Simulation work developed using Opnet 8.0 validates the proposed protection scheme with comparative data for the end-to-end protection versus the proposed Inter-Domain Protection using Domain Boundary Local Bypass Tunnels (IBLBT). A conclusion with proposed additional research conclude this thesis. Appendix A and B provide implementation details and the necessary extensions to existing or proposed protocols and signaling schemes for RSVP-TE.

CHAPTER 2 – STATE OF THE ART

2.1 GENERAL

2.1.1 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 external routing protocols. Research by C. Labovitz et al. [29] 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 restoral”. Although our networks are become more and more resilient there are still frequent network failures which are a cause for concern. “The FCC has reported that network failures in the United States with impact on more than 30,00 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 10 hours”[32]. With aggregation and convergence technologies the problem is made worse and now a single failure may affect millions of users. Strategic planning at Gartner Group suggest at [33] 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 factor behind advances in MPLS protection mechanism.

2.1.2 As briefly discussed at 1.4, 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. 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 rely on the establishments of new paths after the failure event while pre-established mechanism compute and maintain 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 mechanism, one of the first to be 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 between every 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 established on disjoint paths from source to destination. An immediate benefit of end-to-end mechanism is scalability. Reference [11] shows that given a network of N nodes, local repair schemes require $N*(N-1)$ backup paths to protect a network with N 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.

2.2 LOCAL FAST REROUTE

2.2.1 The proposal at [9] and depicted at figure 3 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 document to help differentiate it with the end-to-end Fast Reroute (EFR) described later. In the Local Fast Reroute of figure 3, one-to-one backup LSPs can be established to locally bypass a point of failure.

Figure 3 Local Fast Reroute

2.2.2 A second means of backing up LSPs is to use facility backup by making use of the label stack. Instead of creating a separate LSP for every backed-up LSP, a single LSP is created which serves to backup up a set of LSPs. Such an LSP backing up a set of

primary LSPs is called a bypass tunnel. Two new objects are added to RSVP-TE to accomplish this task. The FAST_REROUTE and DETOUR objects are required to support Local Fast Reroute signaling. Furthermore, the existing SESSION_ATTRIBUTE and RECORD_ROUTE objects are extended to support bandwidth and node protection features. Those new features are necessary to properly signal protection requirements downstream and report back upstream to the source node. With Local Fast Reroute, a link failure at LSR 2-LSR 3 or LSR 3-LSR 4 of figure 3 or a node failure at LSR 3 will result in traffic from LSR 1 and LSR 8 destined for LSR 5 or LSR 9 to be switched at LSR 2 onto the bypass tunnel LSRs 2-6-7-4. Assuming global label assignment at LSR 4 and that the Point of Local Repair LSR 2 has learned the appropriate label by snooping into the RRO objects in the RESV messages, all LSR 2 has to do is insert the expected label and push the bypass tunnel label onto the label stack. For per-interface label assignment, the PLR LSR 2 will have to initiate a path message along the bypass tunnel for every protected LSP under protection of the bypass tunnel. This procedure is necessary for LSR 2 to learn the appropriate labels for the protected LSP that will be recognized by LSR 4. In this case, upon a failure, LSR 2 will replace the protected LSP label with the learned label prior to pushing the bypass tunnel label onto the stack.

2.2.3 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.

2.3 END-TO-END FAST REROUTE

2.3.1 Other methods based on the idea of end-to-end recovery use pre-established backup LSPs computed by the MPLS ingress nodes or PSL. Two flavors of end-to-end schemes are found in the literature. In the first one, such as the end-to-end Fast Reroute (EFR) method, the node immediately upstream from the failure point, effects the switchover to the backup LSP. This technique has the advantage of not requiring fault signaling to the upstream nodes. The end-to-end Fast Reroute method presented at [8] and illustrated at figure 4, trades fast restoration times and simplicity at the expense of comparatively less efficient use of resources. The further away from the source the failure occurs, the less efficient the use of resources. In this scheme a node upon detection of a downstream failure, reroutes the incoming traffic onto pre-established backup LSPs in the reverse direction of the active path. Once the rerouted traffic reaches the ingress node, it is switched to a previously established backup path that is disjoint from the primary path. The rerouting decision is therefore made locally at the point of failure and no signaling is necessary for the reroute decision. A link failure between LSR 2-LSR 3, LSR 3-LSR 4 or

Figure 4 End-to-End Fast Reroute

LSR 4-LSR 5 and/or a node LSR 3, LSR 4 or LSR 5 failure results in the local protection switching from the upstream node to the point of failure. A link LSR 4-LSR 5 failure for example will result in LSR 4 switching traffic onto the backup path label 20. This method has the advantage of reducing the path computation complexity and of not requiring a separate failure notification mechanism to effect switchover. Upon receiving the switched traffic, the ingress node (LER 1) can elect to switch new incoming traffic directly onto the pre-established backup path (label 35). In this case, care must be taken since it is possible that out-of-order packets may be received at the destination. The alternative is to continue sending new traffic onto the failed primary path and let it be switched at the PLR. In this way, no out-of-order traffic will occur at the expense of a less optimal path and inefficient use of network resources.

2.3.2 The main advantage of EFR is its scalability and simplicity while its disadvantages are comparatively long recovery time when compared to LFR due to

propagation delays between the point of failure and the PSL. Another disadvantage is the inefficient use of resources along the protection path.

2.4 REVERSE NOTIFICATION TREE

2.4.1 The second end-to-end scheme is presented at [1] and uses signaling from the point of failure to inform the upstream LSRs that a path has failed. Here a Reverse Notification Tree 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. Figure 5 illustrates the RNT scheme. The RNT is based on the establishment of a Path Switch LSR (PSL) and a Path Merge LSR (PML).

Figure 5 End-to-End protection with Reverse Notification Tree

2.4.2 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. Although the figure shows the ingress LSRs as PSLs, the PSL may be selected at any point along the primary path. The intermediate nodes are responsible to maintain inverse cross connect tables to determine the identity of the upstream nodes that the protected LSPs come from and, create and transmit fault or recovery notifications to each of them when they determine that they are not the PSL for the failed LSP. In addition, the end-to-end RNT scheme introduces a liveness message to detect faults which may not otherwise be detected such as node failures or faults on the reverse link. Several extensions to RSVP-TE are proposed at [1] to enable support of end-to-end path protection using RNT. An explicit Route Protection sub-object is added to the ERO to allow for the identification of the PSL and PML. A Path Protection Object is added to the RSVP path message to configure the protection domain. New Path Protection Error codes are added as well as a Label Request object in the RESV message and Label Object in the Confirmation message to support the RNT.

2.4.3 The main advantages of RNT protection is scalability and efficient use of resources while its disadvantages are long recovery time due to the propagation of failure notification messages.

CHAPTER 3 – PROBLEM STATEMENT

3.1 GENERAL

3.1.1 The MPLS protection mechanism presented in Chapter 2 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. Figure 6 illustrates this problem. Here domains A, B and C all use different protection schemes not to mention that they may also use a different signaling protocol to distribute MPLS

labels. Furthermore, the IRP within domain A only distributes a single route to domain B and C so no alternate path can be computed at the source. Given that in many cases a primary LSP will span more than one administrative domain it is necessary to find a more general protection solution that will solve the protection problem for the multi-domain case. There are several important issues when trying to coordinate LSP and backup LSP establishment across independent administrative domains:

- a) Separate domains may not use the same protection mechanism;
- b) Separate domains may not use the same or interoperable label distribution and signaling protocols;
- c) Separate domains may restrict protection mechanism due to administrative policies;
- d) Fault indication Signaling may not be propagated across independent domains;
- e) The source may not know the level of protection provided in an remote independent domain;
- f) Separate distinct routes may not be known at the source due to the limited external route information distributed by border gateway protocols;

- g) Even in the best case where independent domains cooperate and share the same signaling and protection schemes, latency and sub-optimum resource utilization is a problem;
- h) Long repair latencies may be experienced when end-to-end protection is used due to RNT propagation through multiple domains;
- i) The poor resource utilization of end-to-end Fast Reroute is made worse in the multi-domain case when the primary LSP is switched in the reverse direction back to the source LSR possibly many nodes and domains away; and
- j) It is difficult to provide disjoint paths and full protection at the domain boundaries.

3.1.2 In order to set the stage for the proposed MPLS inter-domain protection solution it is important to examine possible inter-domain connection topologies. The various inter-domain interface connection topologies are presented at figures 7a, 7b, 7c, 7d and 7e. Those topologies are:

- a) Single exit LSR
- b) Single exit LSR, Multi-homed;

- c) Single exit LSR, Multi-homed, Multi-domain;
- d) Dual exit LSRs, and
- e) Dual exit LSRs, Fully Meshed.

Other variations of the above connection topologies are possible but they generally can be reduced to one of those five possibilities. Each of the connection topology requires special attention in the establishment of protection LSPs.

3.2 SINGLE EXIT LSR

3.2.1 In the case of the single exit point LSR (figure 7a), typical in a Stub Network, the source only knows of one way to exit the domain and that is through LSR0. It is therefore impossible for the source to establish an end-to-end backup LSP or even a local repair tunnel completely disjoint from the primary LSP. In the case where local repair is in effect, LSR0 and LSR2 can not be bypassed even if backup redundant links exist between LSR0 and LSR2.

Figure 7a Single Exit LSR

3.2.2 It is easy to see that complete link and node failure protection for the entire path is not possible in this topology. Furthermore even if the source could accept a less than perfect backup path, ie., one that is not completely disjoint from end-to-end, given the inter-domain limitations listed at 3.1.1, it is unlikely that the protection path could be setup successfully.

3.3 SINGLE EXIT LSR, MULTI HOMED

3.3.1 The case of a single exit LSR, multi-homed to the next domain is depicted at figure 7b. As in the previous case, the source LSR is aware of one border router providing connectivity to domain B (LSR0). In this case again, a complete disjoint

Figure 7b Single Exit LSR, Multi-homed

backup path could not be established from source to destination. Requests for node disjoint protection either local repair or end-to-end would fail.

3.4 SINGLE EXIT LSR, MULTI-HOMED, MULTI-DOMAIN

3.4.1 This case depicted at figure 7c is similar to the multi-homed single exit LSR topology other than the fact that a backup LSP at the domain interface could be established through a third domain (domain C) via LSR1. Again, the source LSR is not aware of the secondary route through LSR1 since that secondary external route to the destination will not be advertised by the IRP of LSR0. LSR0 is however a BGP peer to LSR1 and therefore knows of the alternate route to LSR2. As a result, a local protection mechanism at the domain interface with a backup LSP from LSR 0-1-2 could protect

Figure 7c Single Exit LSR, Multi-homed, Multi-domain

against a link LSR 0-LSR 2 failure but will not protect against a node LSR 0 or node LSR2 failure. Complete node and link protection is not possible with this topology.

3.5 DUAL EXIT LSRs

3.5.1 In multiple exit point LSRs (figure 7d) the situation is significantly different than previous single exit point topologies. Here, the source LSR may know the existence of an alternate external route to the destination through a disjoint gateway LSR. This is due to the fact that the IRP of border area routers distribute reachability information independently within their domain. As a result, the source knows of two exit points

providing connectivity to domain B. One route is through border router LSR 0 and the other through border router LSR 1. If the source had routing information into domain B and domain B allowed and supported the protection mechanism and signaling of domain A and fully cooperated, a backup LSP could be established from the source through nodes LSR 1-LSR 3 to the destination.

Figure 7d Dual Exit LSR

3.6 DUAL FULLY MESHED EXIT LSRs

3.6.1 In this case (figure 7e), if LSR 0 is multi-homed to domain B, a backup LSP could be established several ways. Both LSR0 and LSR1 have two links to domain B. Furthermore, the source LSR is aware of two possible exit points to domain B.

Figure 7e Dual Exit LSRs Fully Meshed

Again, if the source LSR had routing information into domain B and domain B allowed and supported the protection mechanism of domain A, a backup LSP could be established from the source through nodes LSR 1-LSR 3 to the destination, hence providing a complete end-to-end disjoint backup path..

3.7 SUMMARY

3.7.1 In the first three topologies presented, it is not possible to provide complete node and link protection for a given protected LSP. For the Dual Exit LSR topologies however, it is possible to provide complete protection. For this reason, this thesis focuses exclusively on the last two Dual Exit LSR topologies.

3.7.2 To solve the inter-domain protection problem it seems obvious that one could simply propose to extend the existing end-to-end intra-domain protection schemes such as EFR and RNT to the inter-domain case. Although this naïve solution may work in some cases, EFR and RNT would not resolve many of the issues listed at paragraph 3.1.1 such as disparate protection schemes between domains, signaling, recovery latency, etc. The application of the naïve extension of end-to-end protection scheme such as RNT will be used in this work to compare the new innovative (IBLBT) protection mechanism presented in this thesis. The IBLBT protection mechanism presented in the next chapter provides a complete and efficient protection solution to the inter-domain protection problem and resolves the issues presented at paragraph 3.1.1.

CHAPTER 4 – PROPOSED SOLUTION

4.1 GENERAL

4.1.1 A complete solution to the inter-domain protection problem can be found if we turn the apparent difficulties in inter-domain LSP setup into advantages. Such is the case for the independence of domains. Accepting the fact that domains will be independent and non-homogeneous leads to the idea of establishing independent 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 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 8 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 path is independent of the other and does not require fault notification beyond its own domain.

Figure 8 Concatenated Primary and Backup LSPs

This innovative solution permits the isolation of the protection mechanism to a single domain or domain boundary. Furthermore, domains can now use independent protection mechanism and signaling schemes and do not need to propagate their 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.

4.1.2 Although the protection domains are isolated, it is still necessary for the domains to exchange some information to let the source know if the required protection can be supported. Although the exact protection mechanism is domain specific, the source needs to know what protection it can offer its client, if any, and if necessary not allow the

connection to be established. Additional signaling is required for this purpose and extensions to existing signaling protocols are proposed at Appendix B.

4.1.3 In summary, the proposed solution to solve the inter-domain MPLS recovery problem is based on the establishment of independent protection mechanism within domains using concatenated primary and backup LSPs, minimal protection signaling between domains and, local repair at the domain boundaries. The 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. Viewed from end-to-end at figure 8, the primary LSP is protected by three or more distinct and independent protection domains merged at their respective boundaries. Those protection domains are the Source Protection Domain, the Domain Interface Protection and the Destination/Transit Protection Domain. In addition to those three protection domains, transit protection domains 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 protection in place.

4.1.4 As discussed earlier this thesis focuses exclusively on the Fully Meshed Dual Exit LSRs and the Dual Exit LSRs topologies. The principles behind the protection scheme are however equally applicable to other topologies capable of supporting complete node and link protection. Figure 9 illustrates the first 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 P1A 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 9. 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. LSR 3 would have no choice but to drop the packets since removing the invalid labels and forwarding the packet as a traditional IP packet may cause loops or it is also possible that the original label was intended to represent a route which cannot be inferred from the IP header. 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. The remainder of this chapter describes the detailed protection setup reasoning and algorithms for each of the protection domains.

4.2 INTER-DOMAIN BOUNDARY LOCAL BYPASS TUNNEL SETUP ALGORITHM

4.2.1 BGP peers are aware of each other's presence through the establishment of direct TCP sessions between BGP speakers. Once the transport protocol session is established, BGP peers exchange an OPEN message. If the OPEN message is accepted, a KEEPALIVE message confirming the OPEN is sent back. Once the OPEN is confirmed, the BGP peer can send UPDATE, KEEPALIVE and NOTIFICATION messages. BGP peers can determine the identity and the location of each other through the lookup of the BGP identifier and Autonomous System (AS) number contained in the messages they exchange. Border LSRs are therefore aware of their external links to border LSRs in different AS and internal links to border LSRs within their own AS. Based on that

information, the IBLBT protection mechanism is established to protect against link and node failures.

4.2.2 The proposed inter-domain boundary local protection scheme is based on the establishment of local bypass tunnels from the Gateway LSR to the PCPSL and from the PGL to the CPSL. The local repair tunnel from the Gateway LSR to the PCPSL provides protection for a link or CPSL failure (entry point LSR) while the local repair tunnel from the PGL to the CPSL protects against a Gateway LSR failure or any other failure on the primary path in the source domain. Given that a Gateway LSR will be a PGL for some LSPs and a Gateway LSR for others, it establishes a local bypass tunnel to both its external BGP peers. In general terms, a full backup mesh is established between external BGP peers.

- a) Case 1 (Fully Meshed Dual Exit LSRs, figure 10), LSR 0 is the Gateway LSR to protected LSP 1 following a path from LSR 0 – LSR 2. LSR 0 would also be the Gateway LSR for a protected path from LSR 0 – LSR 3. LSR 0 is also the PGL to a protected LSP 2 following a path from LSR 1 - LSR 3 or LSR 1 – LSR 2. LSR 1 is a mirror image of LSR 0 for its own paths for which it serves as the Gateway LSR. To cover all cases, LSR 0 establishes a local bypass tunnel to both its external domain BGP peers. A local bypass tunnel is established from LSR 0 to LSR 2 and LSR 0 to LSR 3. LSR 1 executes the same algorithm and establishes backup bypass tunnels to LSR 2 and LSR 3. As result of this setup algorithm, four

separate local bypass tunnels are established to fully protect the domain boundary.

Figure 10 Fully Meshed Dual Exit LSRs

- b) Case 2 (Dual Exit LSRs, figure 11). Again a full mesh of local bypass tunnels is established between external BGP peers. To protect LSP 1 and any other protected LSP on link LSR 0-LSR 2 against a link LSR 0-LSR 2 or an LSR 2 node failure, LSR 0 uses a local bypass tunnel to LSR 3 through LSR 1. LSR 0 also establishes a local bypass tunnel to LSR 3 via LSR 2 to protect traffic on link LSR 1-LSR 3 against a LSR 1 node failure

or any other failure in the source domain. LSR 1 executes the same algorithm resulting in four local bypass tunnels.

Figure 11 Dual Exit LSRs

4.2.3 For the more general case where several external BGP peers exits, it is essential that the CPSL and Gateway LSR select the same PCPSL. Without a common PCPSL it would be impossible to merge the backup traffic from the Gateway LSR with the destination domain backup path originating at the CPSL. Knowledge of the common PCPSL can be obtained by having the Gateway LSR select the PCPSL along the best route to the CPSL. The CPSL would therefore simply select that same PCPSL given that it is on the best reverse route to its Gateway LSR. A more explicit method to identify the

PCPSL is however possible. The alternative method consist of having the Gateway LSR select an external BGP Peer with the alternate best route to the destination (the best route is through the CPSL). That selected secondary BGP Peer becomes the PCPSL for that LSP. The Gateway LSR must then inform the CPSL of that selection. This is necessary so the CPSL will route its destination domain backup LSP through the same PCPSL. In both CR-LDP and RSVP-TE, A new message is required for this identification. The RSVP-TE extensions to support this case are described in Appendix B.

4.3 SOURCE DOMAIN PROTECTION SETUP

4.3.1 Through the reception of external routing information in Link State Advertisement (LSA), the ingress LSR is aware of one or more Gateway LSR providing direct connectivity into the destination or transit domain towards the destination. In OSPF, this information is contained in the Link State Database of all routers within a particular domain. After establishing the primary protected LSP path through the selected Gateway LSR, the Ingress LSR or PSL establishes a backup LSP as follows:

- a) Case 1 (figure 12), the source domain uses an end-to-end protection mechanism and Dual Fully Meshed Border LSRs serve the domain boundary.

Figure 12 Source Domain Protection Setup for Fully Meshed Dual Exit LSRs

The ingress LSR examines its routing tables and selects the best PGL towards the destination domain and establishes a backup LSP to the CPSL. The CPSL terminates the backup LSP and is therefore the merge point for the backup and primary LSPs. The Gateway LSR could not terminate this backup LSP since a requirement exist to provide protection against a Gateway LSR failure. The Gateway LSR must therefore be bypassed by the backup LSP. The Ingress LSR or PSL must be aware of the identity of the CPSL in order to explicitly route the backup path to that node. This information can be provided during the primary path setup using features of the label distribution protocols detailed in Appendix A.

b) Case 2 (figure 13), the source domain uses an end-to-end protection mechanism and Dual Exit LSRs serve the domain boundary. As in the previous case, the ingress LSR examines its routing tables and selects the best Gateway LSR towards the destination. The Ingress LSR establishes the primary path through the selected Gateway LSR using its label distribution protocol. The PSL examines its routing tables and selects the PGL to establish a backup LSP to the CPSL. The backup LSP to the CPSL uses the domain boundary local repair tunnel from LSR 1 to LSR 2 through LSR 3. The CPSL terminates the backup LSP and is therefore the merge point for the backup and primary LSPs.

Figure 13 Source Domain Protection Setup for Dual Exit LSRs

4.4 DESTINATION/TRANSIT DOMAIN PROTECTION SETUP

4.4.1 Within the destination or transit domain and for both Dual Fully Meshed Border LSRs and Dual Border LSRs configurations, the CPSL upon noticing that the primary LSP requires protection, establishes a backup LSP to the destination through the PCPSL as depicted at figure 14. The CPSL selects its PCPSL based on the mechanism described at 4.2.3. In figure 14, the CPSL establishes a backup LSP (B1) to the egress node through the PCPSL (LSR3). The routing of the destination/transit domain backup LSP through the PCPSL is necessary to permit the merging of the local domain boundary backup path with the destination/transit domain backup LSP. Without a common protection node at the domain boundary it would not be possible to achieve full protection while maintaining independent protection mechanism within each domain and domain

Figure 14 Destination/Transit Domain Protection Setup

boundaries. As described earlier, the common PCPSL is the key element to merge the protection domains. The PCPSL merges the backup path originating at LSR 0 with the backup path B1. The details of how this critical label association is performed is explained at Appendix A paragraph 7 for an RSVP implementation.

4.5 SUMMARY

4.5.1 Figures 15 and 16 summarize the complete protection setup in each of the domains for the case of Fully Meshed Dual Border LSRs and Dual Border LSRs respectively. Once established, the protection mechanism offers complete end-to-end link and node failure protection. Although the figures depict local bypass tunnels in one direction only, in practice the bypass tunnels would be established in both directions independently to protect traffic flowing in the reverse direction from domain B to domain A.

4.5.2 For Fully Meshed Dual Exit LSRs, the source establishes the primary LSP (P1A of figure 15) to the destination through its Gateway LSR (LSR0) and CPSL (LSR2) to the egress node (LSR 5). The PSL, which in this case is also the ingress LSR (LSR 9), establishes the source domain backup LSP (B1A) through the PGL (LSR 1) to the CPSL

Figure 15 Pre-Failure Steady State

(LSR 2). This backup LSP uses the pre-established local bypass tunnel at LSR 1 exiting at LSR 2. LSR 0 associates a backup local bypass tunnel label to the primary path and requests a label from LSR 3. LSR 3 merges this backup path label with backup path (B1B). The destination/transit domain protection setup algorithm is executed as described in paragraph 4.4. LSR 2, the CPSL, notices that the primary LSP requires protection and that LSR 3 is the PCPSL for that LSP. LSR2 establishes a backup path for the protected LSP through LSR3 (B1B of figure 15). For the Dual Exit LSR case, the process is the same and is depicted at figure 16.

Figure 16 Pre-Failure Steady State for Dual Exit LSRs

4.6 FAULT DETECTION

4.6.1 To this point a precise mechanism for fault detection upstream from the point of failure has not been discussed. Such a mechanism is of course crucial to the implementation of any recovery scheme at the MPLS layer. LSRs must be able to detect Path Failures (PF), Path Degradation (PD), Link Failures and Degradation (LF) and (LD). The fault detection mechanism must be able to detect unidirectional link faults, downlink faults, uplink faults and bi-directional link and node faults. The MPLS architecture [24] provides for an optional Hello message to supplement existing fault detection mechanism. The proposal at [1] requires a hello message for such fault detection. Upon detecting a failure using this hello message the upstream node from the point of failure generates and forwards a Fault Indication Signal (FIS) to the upstream node towards the PSL. It is suggested that this same fault detection scheme with Hello messages be

introduced at the inter-domain interface to complement existing fault detection mechanism. Because the proposed protection mechanism isolates the various protection domains, edge LSRs at the domain boundaries such as Gateway LSR, CPSL, PCPSL and PGL do not need to transmit FIS upstream. Those LSRs simply use the fault detection of the hello messages to effect local protection switching.

CHAPTER 5 – DETAILED PATH RECOVERY PROCEDURES

5.1 GENERAL

5.1.1 This section describes the detailed procedures to be executed by each participating node in a protection domain. For all scenarios described below, the initial pre-failure steady state is presented at figure 15 and assumes that the domains use independent end-to-end recovery such as RNT.

5.2 GATEWAY LSR PROCEDURES

5.2.1 Refer to figure 17. When a Gateway LSR (LSR 0) detects a failure in its external link, either a link or a LSR 2 node failure, the following protection algorithm is executed.

- a) the Gateway LSR determines what LSPs are affected by the failure and which of those are protected LSPs;
- b) consults the Incoming Label Map (ILM) and the Next Hop Label Forwarding Entry (NHLFE) and extracts the associated local inter-domain bypass tunnel label and the label bound by the PCPSL (LSR 3) for that protected LSP (labels 21 and 12 respectively from figure 17);

Figure 17 Inter-Domain Link or Concatenation PSL Failure Recovery

- c) Remove current labels from the protected LSP and insert new labels. The label stack is now (21,12);
- b) Manipulate the original ERO of the protected LSP to ensure proper handling at the PCPSL. This is necessary since the use of the original ERO would cause the PCPSL to reject the Path messages for the protected LSP since the ERO would include addresses of bypassed nodes;
- c) forward packets onto the bypass tunnel;

- d) Send Path Error Message to the source node indicating a local repair action has taken place;
- e) Do not propagate FIS upstream to cause protection switching at the PSL;
and
- f) Refresh Path State information downstream via the bypass tunnel and RESV State upstream for the protected LSPs.

5.2.2 As a result of those actions, the primary path P1 is switched around the link failure LSR 0-LSR 2 or LSR 2 node failure, to the PCPSL (LSR 3) which merges the backup traffic with the pre-established backup LSP to the egress LSR. The recovery path is LSRs 9-7-0-3-6-5.

Figure 18 Source Domain Failure Recovery

5.3 PROXY GATEWAY PROCEDURES

5.3.1 A failure occurring along the primary path in the source domain as illustrate at figure 18, either a link or Gateway LSR node failure will result in the primary path being switched to the backup path. The switch to the backup path will occur at the PSL. In figure 18, the backup path leads to the CPSL through the PGL (LSR 1) and global label space is used. The PGL pushes its local repair bypass tunnel label (42) onto the stack and forwards the packet to the CPSL. Label 36 was inserted by the PSL to properly associate the traffic at the PML. The recovery path is LSRs 9-8-1-2-4-5.

5.4 CONCATENATION PSL PROCEDURES

5.4.1 A failure point along the path in the destination or transit domain will cause the CPSL to switch to the backup path according to local protection mechanism policies as depicted in figure 19. Upon a link LSR2-LSR4 or node failure at LSR4, the CPSL determines which protected paths are affected by the failure, consults its ILM and NHLFE to extract the backup path labels. The CPSL switches the traffic P1 onto the pre-established backup path P1B through the PCPSL (LSR3). The recovery path is LSRs 9-7-0-2-3-6-5. The CPSL must also manipulate the original ERO before forwarding path messages along the recovery path. The CPSL refreshes Path and RESV messages.

Figure 19 Destination/Transit Domain Failure Recovery

5.5 PROXY CONCATENATION PSL PROCEDURES

5.5.1 When due to a link LSR 0-LSR 2 or LSR 2 node failure the PCPSL detects traffic on a pre-established inter-domain local bypass tunnel (figure 17), the PCPSL executes the following procedures:

- a) Removes the tunnel label and examines the 2nd level label. Consults its ILM to match the incoming label with a backup LSP label;
- b) Extracts the backup LSP label from the table (label 20) and any other associated labels such as label 15 in the case of a bypass tunnel with global address space;
- c) removes previous labels from the stack and inserts the extracted backup LSP label and any other labels necessary for the egress node to properly identify the traffic;
- d) merges the backup traffic with the destination domain backup path; and
- e) Refresh Path messages downstream and RESV messages upstream to the Gateway LSR.

CHAPTER 6 - BANDWIDTH RESERVATION ISSUES

6.1 GENERAL

6.1.1 Along with recovery time and complexity, the amount of bandwidth required to establish a protection scheme is an important element of any performance evaluation matrix. Based on a requirement to provide single node or link failure and referring to figure 10, it is easy to see that for the case of four fully meshed border LSRs, the proposed inter-domain boundary local bypass scheme requires the overall reserved protection bandwidth at the boundary to be equal to the primary path bandwidths. For protected LSP 1 of figure 10 exiting domain A through LSR 0 and entering domain B through LSR 2, the protection scheme reserves an equal amount of bandwidth on the local bypass tunnel from LSR 0-LSR 3. The protection bandwidth on the path LSR 0-LSR 3 is equal to all protected bandwidth on path LSR 0-LSR 2. Furthermore, for the same protected LSP 1, an additional equal amount of backup bandwidth is reserved on the bypass tunnel segment from LSR1-LSR2. This is the protection bandwidth attributed from the source domain backup path. As a result, an initial 200% bandwidth premium for protection is required. This reservation premium is however reduced by half by sharing local bypass tunnel bandwidth with other links carrying protected LSPs. For example, LSP2 and all other protected traffic on the link LSR1-LSR3 are protected by the bypass tunnel LSR1-LSR2 as well. This same bypass tunnel also provides protection against a node 0 failure for traffic on link LSR0-LSR2. In the same way, bypass tunnel LSR0-LSR3 shares protection bandwidth for protected LSPs on link LSR0-LSR2 and for LSPs

on link LSR1-LSR3 against a node 1 failure. As a result of this sharing, the overall additional bandwidth requirement at the domain boundary for the proposed inter-domain boundary protection is 100%.

6.1.2 For cases where the protected LSP is switched from LSR0-LSR3, its backup bandwidth is reserved on link LSR0-LSR2 and LSR1-LSR2 for the backup path originating from the source domain. Again this protection bandwidth is shared with all protected paths on link LSR1-LSR2 and results in a protection premium of 100%. The overall additional bandwidth required for protection is the same as that required for end-to-end protection schemes. Analytically, the reserved backup bandwidth can be calculated as follows:

Let $p_{i,j-k}$ represent the bandwidth of inter-domain protected primary LSP i on path $j-k$ from LSR j to LSR k . The total protected bandwidth on a path $j-k$ is therefore:

$$P_{j-k} = \sum_{i=1}^n p_{i,j-k}$$

Let $b_{j-k,l-m}$ be the reserved backup bandwidth on path $l-m$ for all protected primary LSPs on path $j-k$. When the reserved backup bandwidth for a given protected path is exclusively on a given backup path $l-m$, then $b_{j-k,l-m} = P_{j-k}$ as is the case for the proposed inter-domain protection scheme. The total reserved backup bandwidth on path $l-m$ is b_{l-m} .

There are four possible inter-domain primary paths with the Fully Meshed Dual Exit LSRs at figure 10. They are LSR 0-LSR 2, LSR 0-LSR 3, LSR 1-LSR 2, LSR 1-LSR 3 with the corresponding total reserved backup bandwidth for P_{j-k} equal to the sum of all backup bandwidths for primary path j-k on all backup paths :

$$B_{j-k} = \sum_{l=1}^n \sum_{m=1}^n b_{j-k,l-m}$$

For primary path LSR 4-LSR 5 $B_{4-5} = b_{4-5,4-6} + b_{4-5,3-5}$ similarly

$$B_{4-6} = b_{4-6,4-5} + b_{4-6,3-6}$$

$$B_{3-5} = b_{3-5,3-6} + b_{3-5,4-5}$$

$$B_{3-6} = b_{3-6,3-5} + b_{3-6,4-6}$$

The overall total reserved backup bandwidth at the domain interface for IBLBT is therefore $B_t = B_{4-5} + B_{4-6} + B_{3-6} + B_{3-5}$. We note that for a single link or node protection, the reserved backup bandwidth on path l-m protects two disjoint primary paths. As a result, the bandwidth can be shared and the necessary reserved backup bandwidth on backup path l-m for all primary path j-k is reduced to $b_{l-m} = \text{Max} (b_{j-k,l-m})$.

B_t reduces to :

$$B_t = \sum_{l=1}^n \sum_{m=1}^n b_{l-m} = \sum_{j=1}^n \sum_{k=1}^n P_{j-k} = P_{4-5} + P_{4-6} + P_{3-5} + P_{3-6}$$

6.1.3 For the case of Dual Exit LSR not fully meshed as depicted at figure 11. LSP 1 is protected by an equal amount of reserved backup bandwidth on the inter-domain local bypass tunnel for LSR 2 from LSR 0-LSR 1-LSR 3. LSP 1 is further protected by another equal amount of bandwidth on the LSR 0 bypass tunnel from LSR 1-LSR 3-LSR 2. This backup tunnel is initiated at the source domain PSL and protects against a Gateway LSR (LSR 0) failure. For protected LSPs on link 1-3 such as LSP 2, its backup reserved bandwidth is on the LSR 3 bypass tunnel and LSR 1 bypass tunnel from LSR 0-LSR 2-LSR 3. Again as in the previous case, given protection for a single link or node failure, protection bandwidth can be shared since a single failure will not affect both LSP 1 and LSP 2 simultaneously. The resultant overall protection bandwidth premium is 100%.

6.1.4 Similarly to the analysis for the Fully Meshed case, the required IBLBT backup bandwidth for the Dual Exit LSR case can also be derived. In this case however, only two primary protected paths are possible P_{4-5} and P_{3-6} . Here the total reserved backup bandwidth for P_{j-k} is B_{j-k} and $B_{4-5} = b_{4-5,4-3-6} + b_{4-5,3-6-5}$ and $B_{3-6} = b_{3-6,4-5-6} + b_{3-6,3-4-5}$. We observe that a single reservation of $b_{4-5,4-3-6-5}$ and $b_{3-6,3-4-5-6}$ will protect P_{4-5} and P_{3-6} and therefore $B_t = P_{4-5} + P_{3-6}$.

6.1.5 For both cases, the mechanism to effect the sharing of bandwidth on backup paths is beyond the scope of this work. The contribution at [12] provides a more detailed look at this issue.

CHAPTER 7 - ADVANTAGES

7.1 GENERAL

7.1.1 The innovative IBLBT inter-domain protection mechanism solves the problems of inter-domain protection stated at paragraph 3.1. The advantages of the proposed protection mechanism are:

- a) Permits domain independence. This is a critical advantage as it allows MPLS protection to be established across non-homogeneous domains. With this solution, domains are free to use any signaling and protection mechanism;
- b) Will work for both per-interface and global label distribution cases;
- c) Provided that as a minimum a Dual Exit LSR topology is used, the protection mechanism provides single link and node failure protection from end-to-end;
- d) As a result of the recovery mechanism being contained to a single domain and local at the domain boundary, the recovery is much faster than end-to-end mechanism such as RNT and EFR. In the case where the failure is in the source domain, the proposed protection mechanism performs as well

as RNT or EFR. The further away from the source the failure occurs, the better the proposed scheme performs when compared to other mechanism.

- e) The recovery mechanism makes use of both local repair and end-to-end schemes and therefore combines the advantages of fast local repair with the scalability advantage of end-to-end recovery;
- f) Avoids multi-domain fault notification signaling which causes significant recovery latency and requires failure disclosure to neighboring domains; and
- g) When compared to IP recovery, the proposed scheme avoids potential forwarding blackholes and/or transient forwarding loops when down/up transitions occur on BGP which result in routing flaps.

CHAPTER 8 – SIMULATION

8.1 GENERAL

8.1.1 To verify the potential for the proposed IBLBT solution, three separate models were built. The first one is a simple BGP-4 network consisting of three independent domains. This model was used to gather baseline data on recovery times using traditional layer three inter-domain routing protocols. This data will be used to compare recovery times with the proposed MPLS protection scheme. The second model implements MPLS recovery using an end-to-end protection mechanism. The model was built manually 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 thesis. In contrast to end-to-end recovery, IBLBT isolates recovery to the domain boundary or to an individual domain. The third and final 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.

8.1.2 All models were verified and validated using known operational system configurations and performance characteristics. Results from the new MPLS recovery

mechanism were validated by comparing with expected results based on known recovery characteristics and configured parameters such as link delays.

8.2 BORDER GATEWAY BASELINE MODEL SETUP

8.2.1 Although the model is very simple it does provide a glimpse into very long failure recovery latencies experienced in traditional layer three networks when multiple domains are involved. Figure 20 represents the model built using Opnet 8.0. The links are all DS-3 links running BGP-4 and carrying Email, File Transfers, Database access and Web browsing. The traffic was set at approximately 1.5 Mbps and the links are all 100 km apart. The simulation was run for 15 minutes with a failure point in link 4-2 set for 300 seconds.

Figure 20

8.2.2 For intra-domain recovery, the layer three protocols may be able to recover from failures comparatively rapidly. In the case of OSPF for example, its ability to maintain alternate routes in its Link State Database enables OSPF to quickly re-compute a new route upon failure and introduce this new route in its routing table at the point of local repair. Recovery can therefore take place within a few milliseconds while the complete domain convergence to the new route may take 10s of seconds. For the inter-domain situation studied in the paper, the use of inter-domain routing protocols such as BGP4 makes the recovery latencies much worse. There are several reasons for this as reported in [29]:

- a) Although the adoption of path vector by BGP eliminates the Distance Vector count-to-infinity problem, the path vector exponentially exacerbates the number of possible routing table fluctuations;
- b) The delay in inter-domain route convergence is due almost entirely to the unforeseen interaction of protocol timers with specific vendor implementation decisions; and
- c) The theoretical upper bound on the number of computational states explored during BGP convergence is $O(n!)$, where n is the number of domains in the internet.

8.2.3 It is also important to note that in real networks, BGP does not have any kind of recovery mechanism built into it and BGP processes need to be manually restarted after the recovery of the link. In the Opnet Model, BGP will re-attempt failed connections immediately upon detection of the link recovery. New proposals to improve BGP convergence time have been recently introduced at [30].

8.2.4 In the simulation, upon a link failure between router 4 and router 2, BGP required on average 70 seconds to converge to a new route to router 2 through router 3. Figure 21 presents a graph of received IP packets at node 2 and clearly shows a 73 seconds gap

between the link failure and recovery through route LSR4-LSR3. The link failure was programmed to occur at 300 seconds. Due to the 73 seconds gap before recovery, TCP retransmissions reach the maximum and the connection is reset with an abort issued to the source applications. Table 1 is the extracted routing table entries for router 4 showing that the recovery route for the destination was computed and entered in the routing table at 373 seconds. The recovery times are measured by examining entries in the routing tables and link state databases. The time stamp of those entries and the failure time can be used directly to compute recovery time precisely.

Figure 21 IP packets received at destination Node (router 2)

COMMON ROUTE TABLE snapshot for:

Router name: Enterprise Network.Node_4_Edge_LSR at time: 900.00 seconds

ROUTE TABLE contents:

Dest. Address	Subnet Mask	Next Hop	Metric	Protocol	Insertion Time
192.0.5.0	255.255.255.0	192.0.5.2	150	BGP	70.000
192.0.7.0	255.255.255.0	192.0.7.2	150	BGP	70.000
192.0.6.0	255.255.255.0	192.0.5.1	99	BGP	70.006
192.0.8.0	255.255.255.0	192.0.5.1	99	BGP	70.006
192.0.4.0	255.255.255.0	192.0.4.2	648	OSPF	125.826
192.0.12.0	255.255.255.0	192.0.4.2	648	OSPF	125.826
192.0.2.0	255.255.255.0	192.0.4.2	748	OSPF	125.826
192.0.11.0	255.255.255.0	192.0.11.1	0	OSPF	125.826
192.0.3.0	255.255.255.0	192.0.5.1	98	BGP	373.089
192.0.10.0	255.255.255.0	192.0.5.1	98	BGP	373.089
192.0.1.0	255.255.255.0	192.0.5.1	98	BGP	373.089
192.0.9.0	255.255.255.0	192.0.5.1	98	BGP	373.089

Table 1 Router 4 table entries

8.3 MPLS END-TO-END PROTECTION MODEL

8.3.1 This MPLS model was built to measure recovery time for an end-to-end protection case and is represented at figure 22. As stated earlier, it is recognized that such an inter-domain end-to-end protection mechanism is naïve for the reasons discussed in Chapter 3. However, to obtain comparative data from such a scheme LSPs were configured manually using dynamic LSPs and all nodes share the same signaling and protection mechanism. Traffic was generated using two separate twenty five users Gigabit Ethernet LANs with high resolution video conferencing applications running on 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 22 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.

Figure 22 MPLS End-to-end Protection Model

8.3.2 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);

8.3.3 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.

8.3.4 This models 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.

8.3.5 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 μ sec. This sampling time was selected to provide sufficient granularity into the recovery process while maintaining simulation output files to a manageable size.

8.4 RESULTS FOR END-TO-END MPLS RECOVERY

8.4.1 Link 5-7 failure recovery results. In this failure scenario, LSR 5 detects the failure through the LOS, transmits the LDP notification message to the PSL (LSR 1) which switches traffic to the backup path. In the sample presented, traffic is received from the merge point (LSR 8) until 170.0060 sec reflecting traffic in transit downstream from the failure point. Recovery time is measured to be 11.8 ms. Tables 2 and 3 present a sample LSP output for the primary and backup paths from which the recovery time is calculated.

Based on five replications and a 95% confidence interval, the average recovery time for this scenario was $12.09 \text{ ms} \leq R \leq 14.39 \text{ ms}$ with an average of 13.24 ms.

end-to-end recovery
Link 5-7 failure at 170.0 sec
Primary path (P1) LSP 0-8 output

Time	output bits/sec
170.00570	359999999.9
170.00580	170239999.9
170.00590	119999999.9
170.00600	88639999.9
170.00610	0.000000000000
170.00620	0.000000000000
170.00630	0.000000000000
170.00640	0.000000000000

Table 2

end-to-end recovery
Link 5-7 failure at 170.0 sec
Backup path (B1) LSP 1-8 output

time	output bits/sec
170.01740	0.000000000000
170.01750	0.000000000000
170.01760	0.000000000000
170.01770	0.000000000000
170.01780	359999999.9
170.01790	290239999.9
170.01800	208639999.9
170.01810	119999999.9

Table 3

8.4.2 Link 1-4 failure recovery results: In this failure scenario, LSR 1 detects the failure with the LOS. At this time LSR 1 switches the traffic to the backup path which merges at the egress LSR 8. In the same fashion as in the previous case, the output of the primary and backup paths are monitored and the recovery time is measured to average

7.18 ms with a 95% confidence interval $6.45 \text{ ms} \leq R \leq 7.91 \text{ ms}$. This short recovery time reflects the local failure detection through LOS.

8.4.3 Link 4-5 failure recovery results: In this scenario, LSR 4 detects the failure through LOS and forwards a notification message to the PSL (LSR 1) which switches traffic to the recovery path LSP 1-8. The recovery time R is measured as 12.0 ms on average with $11.56 \text{ ms} \leq R \leq 12.44 \text{ ms}$. The increase in recovery time is due to the 4 ms link delay experienced by the failure notification message propagated from LSR 4 to LSR 1.

8.4.4 Node 5 failure recovery results: In this failure scenario, LSR 4 detects the failure using the keep alive message and after the 30 ms hold off period. LSR 4 sends an LDP notification message to LSR 1 which is the PSL responsible to switch to the backup LSP. Recovery is 46.18 ms on average with $45.88 \text{ ms} \leq R \leq 46.88 \text{ ms}$. The long recovery time is largely characterized by the fault detection time. Figure 23 presents the received IP traffic at the egress LSR (LSR 8) and shows the short drop in received packets at 170 sec. For this graph, packets received are monitored and averaged over a 100 ms period. Figure 24 and 25 present a sample output file at LSR 8 of the primary and backup paths respectively.

Figure 23 Received IP packets at the Egress LSR (LSR 8)

Figure 24 Primary path LSP 0-8 output

Figure 25 Backup Path LSP 1-8 output

8.5 INTER-DOMAIN BOUNDARY BYPASS TUNNEL MODEL SETUP

8.5.1 In the IBLBT model, the primary and backup paths were established following the proposed inter-domain protection algorithms. All LSPs are static LSPs and were built manually into the network model since modification to existing protocols to automate the setup processes is beyond the scope of this work. The goal of this simulation work was simply to obtain basic data supporting the presented benefits of this proposal and not to implement the complete set of algorithms. This work may hopefully encourage further research and development and generate interest to undertake the necessary protocol and process modifications to automate the complete set of algorithms.

8.5.2 As described in previous sections and depicted at figure 26, concatenated LSPs were setup within each domain with backup paths using bypass tunnels established manually as described in section 4. The simulations were run for 130 seconds with failures programmed for 125 seconds. Shorter simulation time are possible with this

Figure 26 MPLS Domain Boundary Local Bypass Tunnel Model

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.

8.6 RESULTS FOR INTER-DOMAIN BOUNDARY LOCAL BYPASS TUNNELS

8.6.1 Link 5-7 failure recovery results: In this scenario border LSR 5 detects the failure through LOS. LSR 5 is also the PSL for this domain and switches the affected traffic to Domain B backup path through LSR 6. The merge point of the primary and backup LSP

is LSR 7 and the output of both primary and backup paths are measured at that node.

Tables 4 and 5 present a sample LSP output files showing the failure point on the primary LSP 5-7 at 125.000 sec and recovery at 125.0088 resulting in a 8.8 ms recovery. This recovery time is characterized mostly by the cumulative 8 ms link delay consisting of 4 ms on link 5-6 and 4 ms on link 6-7. As opposed to the end-to-end model, the failure is isolated to domain B. Figures 27 and 28 present graphically the output of the primary and backup LSPs respectively. The average recovery time was measured to be 8.56 ms with $8.26 \text{ ms} \leq R \leq 8.86 \text{ ms}$.

Local recovery
Link 5-7 failure
Primary path (P1B) LSP 5-7 output

time	output bits/sec
124.99960	290239999.9
124.99970	359999999.9
124.99980	290239999.9
124.99990	50239999.9
125.00000	0.000000000000
125.00010	0.000000000000
125.00020	0.000000000000
125.00030	0.000000000000

Table 4

Local recovery
Link 5-7 failure
backup path (B1B) LSP 5-7 output

time	output bits/sec
125.00840	0.000000000000
125.00850	0.000000000000
125.00860	0.000000000000
125.00870	0.000000000000
125.00880	50239999.9
125.00890	119999999.9
125.00900	119999999.9
125.00910	119999999.9

Table 5

Figure 27 Primary path LSP 5-7 output

Figure 28 Backup LSP 5-7 output

8.6.2 Link 1-4 failure recovery results: In this failure scenario, LSR 1 detects the failure through LOS and switches traffic immediately to the backup path LSP 1-5. Once again LSP output is monitored at the merge point (LSR 5) and recovery time is measured at 4.7 ms with $4.5 \text{ ms} \leq R \leq 4.9 \text{ ms}$ which is largely composed of the 4 ms link 1-3 delay. The failure is isolated to the source domain.

8.6.3 Link 4-5 failure recovery results: In this failure scenario, LSR 4 detects the failure through LOS and switches traffic to the local repair inter-domain bypass tunnel from LSR 4 to LSR 5 which switches the traffic to merge point at LSR 7. Output traffic is monitored at the merge LSR. For a sample LSP output file, the last traffic received on the primary path LSP 5-7 is at 125.0041 seconds while the first traffic received on the recovery path is at 125.0052 seconds. The recovery time is therefore 1.1 ms largely characterized by the inter-domain link 4-6 delay of 0.8 ms. On average the recovery time R was measured to be 1.19 ms and $1.01 \text{ ms} \leq R \leq 1.37 \text{ ms}$.

8.6.4 Node 5 failure recovery results: In this scenario, node 4 detects the failure only after the 30 ms hold off timer expires and the overall recovery is measured at 32.02 ms on average with $31.6 \text{ ms} \leq R \leq 32.44$.

8.7 RESULTS SUMMARY

	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

8.7.1 When compared to BGP recoveries of over 70 seconds demonstrated in this work and recoveries of several minutes widely reported, there is little argument on the recovery benefits of MPLS. Comparing end-to-end recovery with the IBLBT case is not so evident however. The recovery speed benefits of IBLBT over the end-to-end case would have been much more evident had the simulation model include 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.

CHAPTER 9 - CONCLUSION

9.1 GENERAL

9.1.1 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 Wave Division Multiplexing 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 thesis 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 domains and makes no assumption on other domains protection mechanism 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.

9.2 ADVANTAGES

9.2.1 The advantages of the proposed IBLBT protection mechanism are:

- a) Permits complete domain independence;
- b) Will work for both per-interface and global label distribution cases;
- c) Can provide link and node failure protection at the inter-domain interface;
- d) Fast recovery time;
- e) Combines the advantages of fast local repair with the scalability advantage of end-to-end recovery;
- f) Avoids multi-domain fault notification signaling;
- g) Avoids potential forwarding blackholes and/or transient forwarding loops when down/up transitions occur on BGP which result in routing flaps; and
- h) Isolate recovery to a single domain. In the worst case, the proposed protection mechanism is as fast as the intra-domain recovery case.

9.3 DISADVANTAGES

- a) Requires modifications to existing label distribution protocols such as RSVP-TE and CR-LDP;
- b) The requirement to route backup LSPs through Proxy PSL and Proxy Gateway LSRs leads to less than optimum backup routes.

9.4 FURTHER RESEARCH

9.4.1 The next step in this research would be to implement the complete set of inter-domain protection algorithms and protocol modifications to automate the protection setup. With the mechanisms fully automated, it would be possible to test the solution under various inter-domain connection topologies and failure scenarios.

9.4.2 The advantages of the proposed recovery mechanism would be more apparent in a complex network model consisting of numerous independent domains. It is suggested that a more complex network model be developed to further investigate the advantages of this proposal.

9.4.3 Many of the features of the proposed protection mechanism rely on implicit knowledge. It is suggested that additions to existing signaling mechanism be investigated

to provide explicit assignment of recovery elements such as Proxy Gateway LSR and Proxy PSL as proposed in section 6.

9.4.3 The effectiveness and speed of recovery are largely dependent on failure detection and notification mechanism. It is suggested that recovery schemes, including the inter-domain boundary local bypass tunnel proposed in this thesis, be tested using various failure detection and notification techniques to study their impact and optimum settings for recovery performances.

TERMINOLOGY

The terminology used in this paper follows IETF terminology. Some terms are additional terms found in recent IETF proposals while others are completely new terms and are added to help describe some of the processes and algorithms developed as part of this work. The term Domain is used throughout this paper and is synonymous to Autonomous System.

BACKUP PATH [9]: The LSP that is responsible for backing up one protection LSP. A backup path refers to either a detour LSP or a backup tunnel.

BACKUP TUNNEL [9]: The LSP that is used to backup one of the many LSPs in many-to-one backup.

BYPASS TUNNEL [9]: An LSP that is used to protect a set of LSPs passing over a common facility.

CONCATENATION PSL: The border LSR which is the primary LSP entry point into the next domain.

FORWARDING EQUIVALENCE CLASS (FEC) [24]: A group of IP packets which are forwarded in the same manner (e.g., over the same path, with the same forwarding treatment).

GATEWAY LSR: A border LSR which transits a given LSP from one domain to the next.

LABEL [24]: A short fixed length physically contiguous Identifier which is used to identify a FEC, usually of local significance.

LABEL MERGING [24]: The replacement of multiple incoming labels for a particular FEC with a single outgoing label.

LABEL SWAP [24]: The basic forwarding operation consisting of looking up an incoming label to determine the outgoing label, encapsulation, port, and other data handling information.

LABEL SWAPPING [24]: A forwarding paradigm allowing streamlined forwarding of data by using labels to identify classes of data packets which are treated indistinguishably when forwarding.

LABEL STACK [24]: An ordered set of labels.

LABEL SWITCH PATH (LSP) [24]: The path through one or more LSRs at one level of the hierarchy followed by packets in a particular FEC.

MPLS [24]: Multiprotocol Label Switching. The term multi is used because its techniques are applicable to ANY network layer protocol.

PROTECTED LSP [9]: An LSP is said to be protected at a given hop if it has one or multiple associated backup tunnels originating at that hop.

PATH SWITCH LSR (PSL) [1]: An LSR that is the transmitter of both the working path traffic and its corresponding recovery path traffic. The PSL is responsible for the switching of traffic between the working path and the recovery path. The PSL is the origin of the recovery path, but may or may not be the origin of the working path (that is the working path may be transiting the PSL).

PATH MERGE LSR (PML) [1]: An LSR that receives both working path traffic and its corresponding recovery path traffic, and either merges their traffic into a single outgoing path, or, if it is itself the destination, passes the traffic onto the higher layer protocols. The PML is the destination of the recovery path, but may or may not be the destination of the working path. (that is, the working path may be transiting through the PML).

PROXY GATEWAY LSR: A border LSR which transits a backup LSP destined for a neighboring domain. The proxy Gateway LSR is responsible to switch the backup traffic to the concatenation PSL in case of a source domain primary path failure.

PROXY PSL: A border LSR which for a given LSP, transits the destination/transit domain backup LSP originating at the concatenation PSL and the backup LSP from the Gateway LSR for the same protected LSP. For a given protected LSP, the Proxy PSL merges the backup LSP originating from the Gateway LSR and the backup LSP originating from the concatenation PSL.

LIST OF ACRONYMS

BGP	Border Gateway Protocol
CPSL	Concatenation PSL
EFR	End to End Fast Reroute
FEC	Forwarding Equivalence Class
FRR	Fast Reroute
IBLBT	Inter-Domain Boundary Local Bypass Tunnel
ILM	Incoming Label Map
IRP	Interior Routing Protocol
LDP	Label Distribution Protocol
LFR	Local Fast Reroute
LSP	Label Switch Path
LSR	Label Switch Router
MPLS	Multi-Protocol Label Switching
NHLFE	Next Hop Label Forwarding Entry
OSPF	Open Shortest Path First
PCPSL	Proxy Concatenation PSL
PGL	Proxy Gateway LSR
PLR	Point of Local Repair
PML	Path Merge LSR
PSL	Path Switch LSR
RNT	Reverse Notification Tree

REFERENCES

- [1] C. Huang, V. Sharma, S. Makam, K. Owens, "A path Protection/Restoration Mechanism for MPLS Networks," IETF Draft, Work in Progress, <draft-chang-mpls-path-protection-00.txt>, March 2000.
- [2] C. Huang, V. Sharma, S. Makam, K. Owens, B. Mack-Crane, A. Akyol, "Extensions to RSVP-TE for MPLS Path Protection," IETF Draft, Work in Progress, <draft-chang-mpls-rsvp-te-path-protection-ext-01.txt>, November 2000.
- [3] S. Kini, M. Kodialam, T.V. Lakshman, C. Villamizar, "Reservation Protocol with Traffic Engineering Extensions - Extension for Label Switched Path restoration," IETF Draft, Work in progress, <draft-kini-rsvp-lsp-restoration-00.txt>, April 2001.
- [4] D.O. Awduche, L. Berger, D. Gan, T. Li, V. Srinivasan, G. Swallow, "RSVP-TE : Extensions to RSVP for LSP Tunnels", IETF Draft, Work in progress, <draft-ietf-mpls-rsvp-lsp-tunnel-08.txt>, August 2001.
- [5] V. Sharma, B. Crane, S. Makam, K. Owens, C. Huang, F. Hellstrand, J. Weil, L. Anderson, B. Jamoussi, B. Cain, S. Civanlar, A. Chiu, "Framework for MPLS-based Recovery", IETF Draft, Work in progress, <draft-ietf-mpls-recovery-frmwk-02.txt>, March 2001.
- [6] R. Bartos, M. Raman, "A Heuristic Approach to Service Restoration in MPLS Networks," IEEE/ACM Transactions, 2001.
- [7] S. Yoon, H. Lee, D. Choi, Y. Kim, G. Lee, M. Lee, "An Efficient Recovery Mechanism for MPLS-Based Protection LSP," IEEE/ACM Transactions, 2001.
- [8] 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.
- [9] 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.
- [10] I.T. Okumus, J. Hwang, H.A. Mantar, S.J. Chapin, "Inter-Domain LSP Setup Using Bandwidth Management Points," GLOBECOM 2001, Vol 1, page 7-11.
- [11] C. Huang, V. Sharma, S. Makam, K. Owens, B. Mack-Crane, "Building Reliable MPLS Networks Using a Path Protection Mechanism," Carleton University, June 26, 2001.

- [12] S. Kini, M. Kodialam, T.V. Lakshman, S. Sengupta, C. Villamizar, "Shared Backup Label Switched Path Restoration," IETF Draft, Work in progress, <draft-kini-restoration-shared-backup-01.txt>, May 2001.
- [13] R. Goguen, G. Swallow, "RSVP Label Allocation for Backup Tunnels," IETF Draft, Work in Progress, <draft-swallow-rsvp-bypass-label-01.txt>, November 2000.
- [14] P. Ashwood-Smith, A. Banerjee, L. Berger, G. Bernstein, J. Drake, Y. Fan, K. Kompella, E. Mannie, J.P. Lang, B. Rajagopalan, Y. Rekhter, D. Saha, V. Sharma, G. Swallow, Z. Bo Tang, "Generalized MPLS Signalling – RSVP-TE Extensions," IETF Draft, <draft-ietf-mpls-generalized-rsvp-te-00-txt>, November 2000.
- [15] A. Orda, "Routing with End-to-End QoS Guarantees in Broadband Network," IEEE/ACM Transactions on Networking, Vol 7, No 3, June 1999.
- [16] R. Doverspike, J. Yates, "Challenges for MPLS in Optical Network Restoration," IEEE Communications Magazine, Feb 2001.
- [17] A. Banerjee, "Fault Recovery for Guaranteed Performance Communications Connections," IEEE/ACM Transaction on Networking, Vol 7, No 5, Oct 1999.
- [18] S. Shenker, L. Breslau, "Two Issues in Reservation Establishment," SIGCOMM 95, Cambridge, MA.
- [19] C. Assi, Y. Ye, A. Shami, S. Dixit, I. Habib, M.A. Ali, "On the Merit of IP/MPLS Protection/Restoration in IP over WDM networks," IEEE/ACM Transaction on Networking, 2001.
- [20] S. Han, K.G. Shin, "Fast Restoration of Real-Time Communication Service from Component Failures in Multi-Hop Networks," SIGCOMM 97, Cannes, France.
- [21] R. Bartos, M. Raman, "New Approaches to Service Restoration in MPLS-based Networks," IEEE/ACM Transaction on Networking, 2001.
- [22] T.H. Oh, T.M. Chen, J.L. Kennington, "Fault Restoration and Spare Capacity Allocation with QoS Constraints for MPLS Networks," IEEE/ACM Transaction on Networking, 2000.
- [23] F. Hellstrand, L. Anderson, "Extensions to CR-LDP and RSVP-TE for setup of pre-established recovery tunnels, IETF Draft, <draft-hellstrand-mpls-recovery-merge-01.txt> November 2000.
- [24] E. Rosen, A. Viswanathan, R. Callon, "Multiprotocol Label Switching Architecture," IETF RFC 3031, January 2001.

- [25] R. Braden, L. Zhang, S. Berson, S. Herzog, S. Jamin, "Resource ReSerVation Protocol (RSVP) – Version 1 Functional Specification," IETF RFC 2205, September 1997.
- [26] Y. Rekhter, T.J. Watson, T. Li, "A Border Gateway Protocol 4 (BGP-4)," IETF RFC 1771, March 1995.
- [27] Y. Rekhter, T.J. Watson, P. Cross, "Application of the Border Gateway Protocol in the Internet," IETF RFC 1772, March 1995.
- [28] J. Moy, "OSPF Version 2," IETF RFC 2328, April 1998.
- [29] C. Labovits, A. Ahuja, A. Bose, F. Jahanian "Delayed Internet Routing Convergence," IEEE/ACM Transactions on Networkin, Vol 9 No 3, June 2001.
- [30] S.R. Sangli, Y. Rekhter, R. Fernando, J.G. Scudder, E. Chen, "Graceful Restart Mechanism for BGP," IETF Internet Draft, draft-ietf-idr-restart-05.txt, Dec 2002.
- [31] Y. Rekhter, R. Aggarwal, "Graceful Restart Mechanism for BGP with MPLS," IETF Internet Draft, draft-ietf-mpls-bgp-mpls-restart-00.txt, Jan 2002
- [32] P. Demeester, T. Wu, N. Yoshikai, "Survivable Communications Networks," IEEE Communications, Aug 1999, Vol. 37 No 8.
- [33] B. Hafner, J. Pultz, "Network Failures: Be afraid, be very afraid," Gartner Group, note SPA-09-1285, 15 September 1999.

APPENDIX A – IMPLEMENTATION WITH RSVP-TE

1 GENERAL

1.1 To this point, several algorithms to setup the source, boundary, destination/transit domain protection have been described. In order for those algorithms to work, it is essential that the executing nodes have the necessary information to first be aware of their responsibilities and also to know what algorithms to execute and when to execute them. This Appendix presents the information requirements and implementation details for the case where the label distribution protocol is RSVP-TE. The role and requirements for each protection element such as the source or PSL, the Gateway LSR, PGL, CPSL and PCPSL will be examined. The example used for the description will be limited to the Fully Meshed Dual Exit LSRs case where both independent domains use an end-to-end protection mechanism such as RNT. Appendix B concludes with a description of the required protocol extensions and modifications.

2 SOURCE

2.1 The source/ingress node, must have the following information:

- a) Address of all Gateway LSRs giving access to the destination or transit domain. This information is obtained through existing BGP and IRP protocols. External Route information is flooded to all routers within a

domain. The source LSR extracts the Router ID, from the common header of an OSPF packet. In the case of routes external to the source domain, the OSPF packet is a type 5 link state packet. The source LSR also must extract the external route tag of the type 5 link state packet. The external route tag must contain the AS number of the next AS along the path to the destination. This information provided via OSPF LSA messages is the building block of the Link State Database;

- b) Address of the preferred Gateway LSR. This is basic routing information found in the router ID from type 5 link state packet and the metric information; and
- c) Knowledge that adequate protection as per requested QoS will be provided in the destination and transit domains. This information is not currently available. Proposed extensions to RSVP SESSION_ATTRIBUTE and Path Error message Error Values are described in Appendix B and resolve this issue.

3 PSL

3.1 The PSL must have the following information:

- a) For a given protected LSP, the PSL must have knowledge of the identity of the PML. If no source domain PML is identified in the Explicit Route Protection sub-object (ie: protection is required from end-to-end across domains) and if a PGL exist, the PML is the CPSL. The address of the CPSL can be extracted from the RRO object in the RESV messages originating at the destination; and
- b) The PSL must maintain all necessary protection setup and label associations for the protected and backup LSPs.

4 GATEWAY LSR

4.1 The Gateway LSR must have the following information:

- a) This LSR must know its inter-domain and intra-domain BGP Peers and which ones provide access to common domains. Existing BGP4 protocol provides this information;
- b) Knowledge that it is the Gateway LSR and what CPSL to use for a given protected LSP. It acquires this knowledge by determining that it is the exit LSR for that LSP and consulting its routing table for the next best hop toward the destination;

- c) Knowledge of the PCPSL for the protected LSPs. The PCPSL is the next domain border LSR on the best route to the concatenation PSL. The PCPSL may also be the next domain entry LSR on the best route towards the destination. When this later technique is used the PCPSL must be explicitly identified using a new RSVP-TE Path message object defined in Appendix B.
- d) The Gateway LSR maintains in its Incoming Label Map (ILM) binding information for inter-domain local bypass tunnels for all protected LSPs. The ILM will therefore contain at least two Next Hop Label Forwarding Entries (NHLFE) one for the primary LSP leading to the CPSL and one for the inter-domain local repair bypass tunnel leading to the PCPSL. The Gateway LSR must also obtain a new label from the PCPSL for every protected LSP. This other label will be used in the stack with the bypass tunnel label to unambiguously bind, at the Proxy PSL, the backup traffic from the Gateway LSR to the backup LSP from the CPSL.

5 CONCATENATION PSL

5.1 The CPSL must have the following information:

- a) This LSR must know its inter-domain and intra-domain BGP Peers. Existing BGP4 protocol provides this information;

- b) Knowledge that it is the concatenation PSL for a given LSP. The border LSR examines the previous hop (PHOP) object of each path message it receives. Based on that, the border LSR can determine that the IP address of the PHOP is a BGP peer in the neighboring domain. It can therefore infer that it is the CPSL for that LSP;
- c) For protected LSPs, the RRO including labels is included in the RESV messages flowing upstream to the source node. The CPSL must maintain this information for protected LSPs. In the case of global label space, the CPSL may use the label information found in the RRO object to properly assign a label to the backup LSP prior to pushing the bypass tunnel label onto the stack; and
- d) Know what internal border LSR is the PCPSL for a given protected LSP. The CPSL gets this information by looking at its BGP4 Adjacency Routing Information Base In (Adj-RIBs-In) and determining which local BGP peer LSRs provides connectivity to the source domain and which of those provides the best route to the source domain Gateway LSR for that protected LSP. As previously described, an alternative explicit method is possible where the Gateway LSR assigns the PCPSL and informs the CPSL of the selected LSR ID. This can be done using the new proposed PCPSL ID sub-object describe in Appendix B.

- e) Knowledge that it is the PML for a given LSP. It obtains this knowledge using the backup path signaling techniques described in paragraph 8.

6 PROXY GATEWAY LSR:

6.1 The PGL must have the following information:

- a) This LSR must know its inter-domain and intra-domain BGP Peers.
Existing BGP4 protocol provides this information; and
- b) Maintain label binding information for all its local, domain boundary bypass tunnels.

7 PROXY CONCATENATION PSL

7.1 The PCPSL must have the following information:

- a) For a given protected LSP, knowledge of the label binding and association information between the backup LSP originating at the Gateway LSR and the backup LSP originating at the CPSL. The PCPSL is in effect the merge point for those two backup LSPs

providing protection for the same Protected LSP. The PCPSL merges the two backup LSPs using the Session and Sender Template Objects which are the same as that of the protected LSP. The objects may differ in sender IP addresses. In this case, the extended tunnel ID in the backup path message will correspond to the source IP address and match the sender IP address in the protected LSP path message.

8 SIGNALING FOR BACKUP PATHS

8.1 Reference [9] presents a number of objectives for a signaling solutions in the setup of backup Paths. Those objectives are:

- a) Unambiguously and uniquely identify backup paths;
- b) Unambiguously associate protected LSPs with their backup paths;
- c) Work with both global and non-global label spaces;
- d) Allow for merging of backup paths; and
- e) Maintain RSVP state during and after fail-over.

8.2 With RSVP-TE several methods are possible to identify and associate protected LSPs with their respective backup paths. LSP tunnels can be identified by a combination of the SESSION and SENDER_TEMPLATE objects. Backup paths can be identified using a Path Message specific approach where the backup paths use the same SESSION and SENDER_TEMPLATE objects as the ones used in the protected LSP. With this approach, some other information must be provided to allow the LSRs to differentiate the backup paths from the protected LSPs. For end-to-end protection with RNT, flags are used for this task while the presence of other objects, such as the presence of Detour objects, can be used for local repair schemes. An alternate way to identify backup paths is to copy the SESSION object and LSP_ID from the protected LSP but set the IPv4 tunnel sender address to the address of the PLR or PSL. In this case, the Extended Tunnel ID of the Session Object must be set to the source IP address.

8.3 The merging operation performed on recovery path is different than the usual RSVP merge. This is due to the fact that in common merge operation an update of resource utilization is made. In the case of recovery paths merging with primary paths, this update is not necessary since traffic will be coming from either the primary or recovery path but never both at the same time. It is therefore necessary in the inter-domain bypass tunnel mechanism to differential primary and recovery/backup paths. The proposal at [23] suggest the addition of a new flag in the Session Attribute to identify a recovery path.

0x09 Recovery Path

This flag indicates that the tunnel is setup as a pre-established recovery tunnel and should be physically merged with the primary LSP at the node where they share the same session object.

8.4 With the proposed inter-domain boundary local bypass tunnel, it is possible that the merge operation is between two recovery paths. As a result, the merge operation should be conducted for all paths which share the same session objects whether primary or backup.

9 PROTECTION SETUP WITH RSVP-TE

9.1 With the information described above and referring to figure 15, the ingress LSR (LSR 9) inserts an Explicit Route Protection sub-object [2] in the path message identifying the desired PSL and PML for this protected LSP. Assuming that the primary path requires protection from end-to-end, the ingress LSR inserts the destination address as the router ID of the PML. The PSL identifies itself from the path message Explicit Route Protection sub-object. The PSL inserts a Path Protection Object to configure the source protection domain. At this point it is likely that a bypass tunnel to the CPSL already exist. In that case, and provided global address space is used, the PSL only needs to extract the expected label at the CPSL from the RESV RRO and make the necessary entry into its forwarding database to associate the protected LSP label with its bypass

tunnel and tunnel end point labels. For per interface labeling, the PSL must, for every protected LSP, send a path message along the bypass tunnel to the CPSL requesting a label for backup LSP.

9.2 If a bypass backup tunnel from the PSL to the CPSL has not been previously established, the PSL establishes one using the CPSL address extracted from the RRO object. The resulting backup LSP to the CPSL through the Gateway LSR (LSR 1) is B1A of figure 15. The CPSL removes Explicit Route Protection sub-objects and Path Protection objects from path messages. Within Domain B, the CPSL inserts a new Path Protection Object to configure the protection within Domain B and uses an Explicit Route Object to route the backup path through the PCPSL (LSR 3).

APPENDIX B - MODIFICATIONS AND EXTENSIONS

1 GENERAL

1.1 The information requirements stated at appendix A are key elements in the establishment and maintenance of the proposed protection mechanism. The generation and distribution of this information requires the modification or extension of existing signaling elements in RSVP-TE. This appendix describes those modifications and extensions.

2 NOTIFICATION OF LOCAL REPAIR

2.1 In general, a recovery mechanism needs to inform the source node that the protected path has been switched to a backup path. This is necessary to give the source node a chance to re-compute, if possible, a new optimum path. Such notification mechanism are possible in RSVP-TE by requiring the Point of Local Repair (PLR) node to send a Path Error Message with error code "NOTIFY" (Error code =25) and an error value field of ss00 cccc cccc cccc where ss=00 and the sub-code = 3 to indicate "Tunnel locally repaired". In the case of end-to-end protection schemes, the Fault Notification Signaling (FIS) provides this notification to the PSL within a domain. For some faults occurring at the domain boundary, the Gateway LSR is the PLR and is required to send a Path Error Message to the source.

3 ROUTE OF ACTIVE PATH

3.1 Modifications to the existing proposal presented at [3] are required. The new proposed extension is necessary for the global label assignment case to provide the PCPSL with the expected label at the bypass tunnel egress point. The Route of Active Path C-Type = 2 of reference [3] needs to be extended to also include the labels assigned along the primary path much in the same way as is provided in the RRO. Label Sub-objects of the Record Route Objects can simply be copied as sub-objects on the proposed backup LSP Information Object.

4 PROTECTION LEVEL SIGNALING

4.1 To provide QoS guarantees to its customers, service providers must know, among other things, the protection mechanism provided for the entire protected LSP and not just for the source domain segment of the path. In this way, if a path can not be protected adequately to meet the customer requirements, the ingress node may refuse the connection or try again at another time. As discussed previously, independent domains do not necessarily support identical protection schemes and it is therefore essential that a signaling scheme be developed to provide the necessary information to the ingress LSR. To establish the protection path, the source must inform the LSR along the paths, both primary and backup, of the requirement and type of protection requested. LSRs must in turn inform the source of any problems providing the requested protection level.

4.2 Signaling is required in the Path message of RSVP-TE to be carried along the primary path to inform the primary path LSRs of the requested protection type, level and attributes. Although much work has been done in the establishment of individual protection schemes, little consideration was given to the possibility that a protected LSP may be protected using several different independent mechanisms. To address this shortcoming, it is proposed that the Session Attribute Object of RSVP-TE [11] be modified to signal the requested protection level as follows:

SESSION_ATTRIBUTE Class = 207, LSP_TUNNEL C-Type = 7

Current Flags:

0x01 Local Protection desired

4.3 This flag permits transit routers to use a local repair mechanism which may result in violation of the explicit route object. When a fault is detected on an adjacent downstream link or node, a transit node can reroute traffic for fast service restoration.

0x02 Label recording desired

4.4 This flag indicates that label information should be included when doing a route record.

0x04 SE Style desired

4.5 This flag indicates that the tunnel ingress node may choose to reroute this tunnel without tearing it down. A tunnel egress node **SHOULD** use the SE Style when responding with a RESV message.

New Flags:

0x20 Protection Level 1

4.6 This flag indicates that the protected LSP requires guaranteed service restoration and that all links and nodes in all protection domains shall be protected by at least one disjoint backup LSP. Furthermore, this flag indicates that the backup path shall provide the same level of service as the primary LSP.

0x22 Protection Level 2

4.7 This flag indicates that the protected LSP requires guaranteed service restoration and that all links and nodes in the source, destination and transit domains shall be protected by at least one disjoint backup LSP. The primary and backup LSP may share a common border LSR and/or link at the domain interface. Furthermore, this flag indicates that the backup path shall provide the same level of service as the primary LSP.

0x24 Protection Level 3

4.8 This flag indicates that the subject LSP requires protection and that all links and nodes throughout the path shall be protected by at least one disjoint backup LSP whenever possible. Furthermore, this flag indicates that the backup path shall provide the same level of service as the primary LSP.

0x26 Protection Level 4

4.9 This flag indicates that the subject LSP requires protection and that all links and nodes throughout the path shall be protected by at least one disjoint backup LSP whenever possible. Furthermore, this flag indicates that the backup LSP may provide a reduced level of service when compared to the primary LSP.

0x28 Protection Level 5

4.10 This flag indicates that the subject LSP requires protection but a best effort level of protection is acceptable. Furthermore, this flag indicates that the backup LSP may provide a reduced level of service when compared to the primary LSP.

4.11 Not only is it necessary that the ingress LSR communicate its customer's required protection level but the downstream nodes must communicate upstream to the ingress LSR when a requested protection level can not be satisfied. To accomplish this task, when the required level of protection is not supported by a downstream node, the node making that assessment shall send a Path Error message to the source. New path error codes in the RSVP Path Error messages are necessary to inform the source that a downstream LSR does not support the protection level requested. The following additional error values are proposed for Error Code 24 "Routing Problem":

Value	Error:
14	Protection mechanism not supported (extension proposed at 2);

New proposed error values

Value	Error:
15	Protection Level not supported due to local policy;
16	Protection Level not supported; and
17	No protection supported.

5 EXTERNAL ROUTE INFORMATION

5.1 It is important to note that if more than one border LSR provides connectivity to a common external domain, the internal routing protocol is aware of those LSRs. In the case of OSPF, all routers within a domain obtain this information through Type 4 and Type 5 Link State Packets flooded into the domain. In order for sufficient information be provided in the Type 5 Link State Packet, the four octets external route tag must be used to carry the AS number of the next AS hop on the path to the advertised destination. With this information, the source LSR knows all border LSRs providing access to a common external AS. The use of the Multi-Exit Disc (BGP-4) attribute may be used by a BGP speaker to discriminate among multiple advertised exit points.

6 PROXY CONCATENATION PSL IDENTIFICATION

6.1 It was identified at section 4.3.3 that an alternate method of identifying a PCPSL is for the Gateway LSR to select a neighboring border LSR on the alternate best route towards the destination. That selected neighboring border LSR in the destination/transit domain is the elected PCPSL for the given protected LSP. The Gateway LSR must then inform the CPSL of this selection. The PROXY PSL ID is identified using the proposed Explicit Route Protection sub-object at [2] with the following modifications:

As stated at [2] this subobject could be strict or loose. The suggested modification is to increase P from 1 to 3 bits. The P bits would now be:

000 For PSL

001 For PML

010 For Proxy PSL