Memòria del treball de Recerca
Research Project

MPLS dynamic multilevel protection

Protecció dinàmica i multinivell d'entorns MPLS (Multiprotocol Label Switching)

Presentat per (by):

Eusebi Calle Ortega

Programa de doctorat (PhD program):

IITAP (Informàtica Industrial / Tecnologies avançades de producció)

Departament d'Electrònica, Informàtica i Automàtica.

Universitat de Girona

Girona, 26 de Juliol de 2001

Acknowledgements

First of all I would like to thank those members who evaluate my work. I am also grateful to my thesis director Dr Josep Lluís Marzo and to my thesis director Dr Teo Jové. Finally I want to thank all the rest of our research group (especially to Dr Ramon Fabregat) in our department at the University of Girona.

Index

Abstract

In this research project, a survey of MPLS (Multiprotocol Label Switching) protection methods and their utilization in combination with on-line routing methods is introduced. Topics such as MPLS, fault management methods and QoS routing are reviewed in this research report. Fault management methods usually pre-establish backup paths to recover the traffic after a failure. In addition, MPLS allows us the creation of different backup types, likewise MPLS is a suitable method to support traffic engineered networks. An introduction of several LSP (Label Switch Paths) backup types and their advantages and disadvantages are introduced. The creation of an LSP involves a routing phase, which should include QoS aspects. In a similar way, to achieve a reliable network, the establishment of LSP backups must also be routed by a QoS routing method. When LSP creation requests arrive one by one (a dynamic network scenario), on-line routing methods are applied. A review of QoS routing and several MPLS on-line routing proposals are introduced. The relationship between MPLS fault management and QoS on-line routing methods is unavoidable, in particular during the creation of LSP backups. Both aspects are discussed in this report. A proposal of an MPLS dynamic multilevel protection, which includes MPLS protection and on-line routing algorithms, is introduced.

1 Chapter

Introduction and background

1.1- Multiprotocol Label Switching

1.1.1.- Introduction

Multiprotocol Label Switching emerged from the evolution of routing/forwarding protocols. MPLS delivers a solution that integrates the control of Level 3 routing with the simplicity of Label 2 switching. Basically MPLS contributes to the separation of control and forwarding components and the Label-swapping forwarding algorithm [ROSE-98]. Figure 1 shows the separation of the control plane and the forwarding label.

The control component (management engine) has two main functions: Path discovery (routing), that involves creating the routing tables, and the signaling function (to signal a path routed). The routing protocol exchange information with other routers to build and maintain a routing table, using standard level 3 routing protocols (OSPF or BGP-4, see [MOY-98, [REKH-95]). The forwarding table is maintained from the control engine and is distributed along network nodes from a signaling protocol (Reservation Protocol RSVP or Label Distribution Protocol LDP).

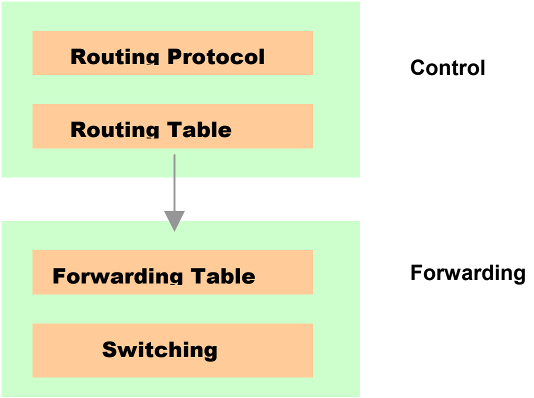


Figure 1 Control and forwarding components

The forwarding component is based on a label-swapping forwarding algorithm (the same algorithm used to forward packets in ATM and Frame Relay switches). Signaling protocol and label distribution allows the creation of the Label Swapping Paths (LSP) similar to ATM Virtual Paths (VPI/VCI).

A label is a short fixed-length value carried in the packet's header to identify a Forwarding Equivalence Class (FEC). An FEC is a set of packets that are forwarded over the same path.

1.1.2 MPLS Header

The 32 bits MPLS header contains the following fields:

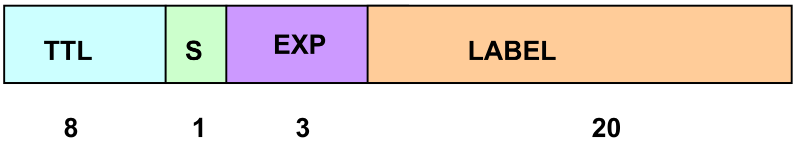


Figure 2. MPLS Header

The label field (20 bits) carries the actual value of the MPLS header.
The EXPerimental field (3 bits) for QoS provisioning.
The Stack field (1 bit) supports a hierarchical label stack.
The Time To Live field (8 bits) provides conventional IP TTL functionality.

1.1.3.- MPLS architecture

Multiprotocol Label Switching is thought of as a heterogeneous protocol where different network components are able to be be founded. For example in an MPLS backbone could coexist with

IP routers without MPLS capabilities or NIF (native forwarding) routers, with ATM-MPLS switches and MPLS routers called LSR (Label-Based Switch Routers). These last routers are given a different name if they are located in the MPLS backbone (where they are called core routers) or at the edge of the backbone (where they are called Label Edge Routers LERs). LER routers are ingress routers and egress routers, depending on whether or not they are a source node or the end node. (Figure 3 shows an MPLS backbone, formed with an ingress node, an egress node and all intermediate LSR nodes).

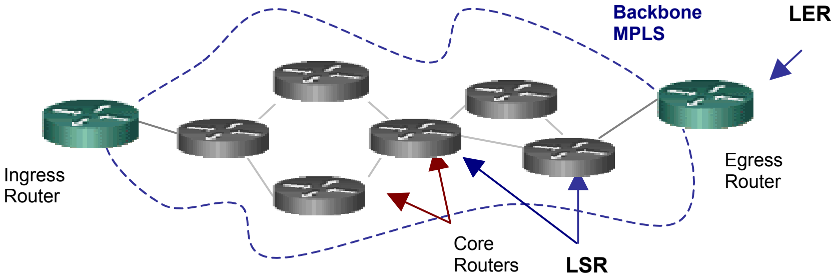
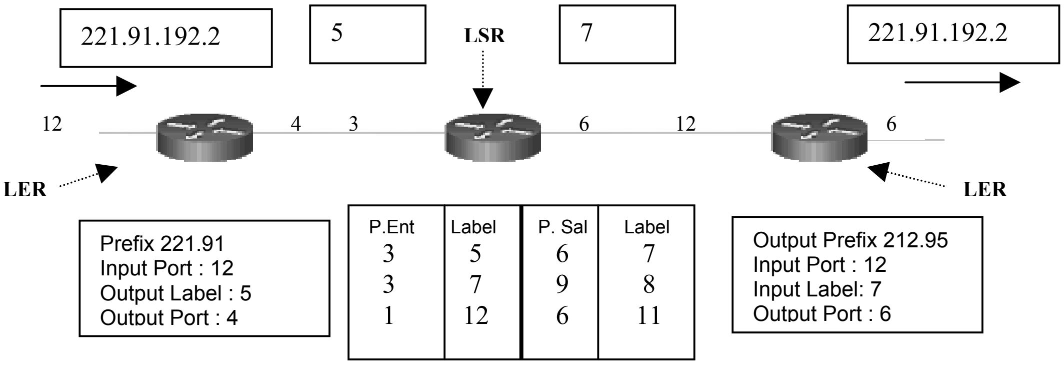


Figure 3. MPLS architecture

1.1.4.- MPLS operation

Using a conventional routing protocol and a signaling protocol usually a Label Distribution Protocol (LDP) or a Reservation Protocol (RSVP), Label Switch Routers build forwarding tables and distribute their labels into them, creating a MPLS path called LSP (Label Switch Path). An Ingress node computes "edge LSR function", which means that it applies an initial label to an IP ingress packet, after examining the IP header. Once a time label is assigned the next LSR can only execute the forwarding function by using this label. LSRs compute forwarding functions using label swapping paradigms (exchange labels at each LSR). At the end of the LSP the egress node computes the reverse function (changing the MPLS label for an IP direction). Figure 4 shows an example of an MPLS operation.



Prefix 221.91
Input Port : 12
Output Label : 5
Output Port : 4

P.Ent	Label	P. Sal	Label
3	5	6	7
3	7	9	8
1	12	6	11

Output Prefix 212.95
Input Port : 12
Input Label: 7
Output Port : 6

Figure 4. MPLS operation

MPLS allows hierarchical labels to be supported as a LIFO label stack [ROSE-00]. A packet is always processed based on the top label and regardless of the other labels that may be below it. In a label stack, the label at the bottom of the stack is called the level 1 label, and labels above it are numbered consecutively up to the level n level. After the top label is processed, a router may pop or push the label stack.

1.1.5.-MPLS Applications

In this section a review of main MPLS applications is introduced. None of these applications are the objective of this research project, but this overview allows a better understanding of the MPLS technology principles.

1.1.5.1.- IP over ATM

MPLS directly provides IP services over ATM switches. Both, routing IP and signaling software could be integrated in ATM switches. MPLS labels are directly mapped in VCI/VPI ATM fields.

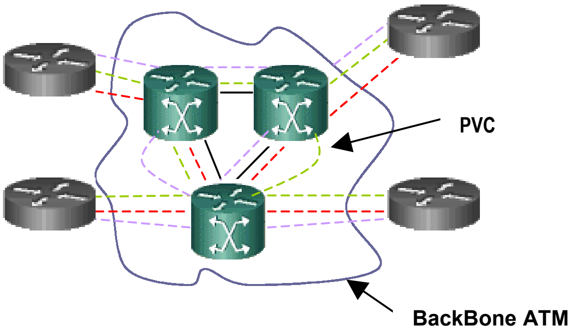


Figure 5. Overlay model (IP over ATM)

Basically, MPLS respects other IPs over ATM mechanisms, MPLS offers more scalability and simplicity. IP over ATM mechanisms, such as MPOA (Multiprotocol Over ATM), involves creating permanent connections (PVC) between edge ATM backbone components. This means a scalability matter, because the network grows exponentially n*(n-1) to create a full mesh to all nodes (overlay model). Moreover, other problems arise, for instance IP over ATM cell transport adds an overhead (about 20 %) and overlay model involves managing two different schemes (ATM and IP). (Figure 5 show the overlay model)

With MPLS we do not have to administrate two separate architectures, changing IP directions and ATM routing tables. MPLS separate routing and forwarding components, so an ATM is only responsible for transporting cells (see [DAVI-01] for more details).

An added benefit of MPLS implementation over an existing ATM network is that it is not required that every device in the MPLS domain should be an LSR. MPLS can be implemented in the same network that is also simultaneously operating standard Layer 2 protocols, known also as “ships in the night”. Neither does it impose additional configuration on a non-MPLS, allowing the network more freedom when designing and migrating to MPLS from an existing network infrastructure. Moreover, ATM switches that currently offer multiservice features can continue to provide ATM services as usual whilst being migrated to an MPLS environment.

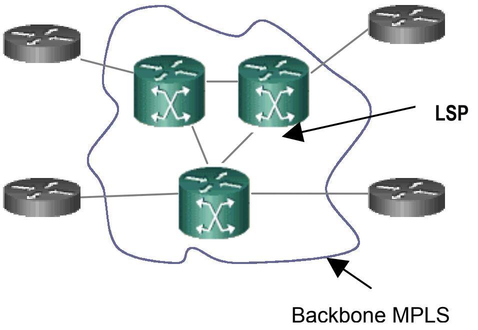


Figure 6. MPLS model

1.1.5.2 Traffic Engineering

Traffic Engineering (TE) involves several aspects related to capacity management and traffic management [AWDU-00]. Capacity management aspects include capacity planning, routing control, and resource management. Network resource management is an important aspect including link bandwidth management, buffer space control and computation resource utilization. On the other hand, traffic management includes: nodal traffic control (such as traffic conditioning, queue management, scheduling) and other functions that regulate traffic flow through the network or that arbitrate access to network resources between different packets or between different traffic streams.

MPLS networks allow us explicit routing of packets by putting labels on them, which can be used to forward packets along specific paths. This encapsulation can be implemented at ingress routers (routers at the edge of an MPLS backbone) to achieve certain QoS requirements. This aggregation (mapping traffic into “forwarding equivalence traffic” FECs) added to the explicit routing property, allows an MPLS to be a useful tool to develop Traffic Engineering networks.

TE-MPLS advantages:

Several proposals to develop an MPLS Traffic Engineering framework are proposed in the literature, such as [AWDU-99c], [XIAO-00], [SWAL-99] and [GHAN-99]. In [AWDU-99] MPLS traffic-engineering advantages are defined:

1.- Explicit label switched paths that are not constrained by the destination based forwarding paradigm can easily be created through manual administrative action or through automated action by the underlying protocols.
2.- LSPs can potentially be efficiently maintained.
3.- Traffic trunks can be implemented and mapped onto LSPs.
4.- A set of attributes can be associated with traffic trunks which modulate their behavioral characteristics.
5.- A set of attributes can be associated with resources which constrain the placement of LSPs and traffic trunks across them.
6.- MPLS allows for both traffic aggregation and desegregation whereas classical destination only based IP forwarding permits only aggregation.
7.- It is relatively easy to integrate a "constraint-based routing" framework with MPLS.
8.- A good implementation of MPLS can offer significantly lower overhead than competing alternatives for Traffic Engineering.

Traffic trunks

A traffic trunk is an aggregation of traffic flows of the same class which are placed inside a Label Switched Path [AWDU-99]. Essentially, a traffic trunk is an abstract representation of traffic to which specific characteristics can be associated. It is useful to view traffic trunks as objects that can be routed; that is, the path through which a traffic trunk traverses can be changed. In this respect, traffic trunks are similar to virtual circuits in ATM and Frame Relay networks. It is important, however, to emphasize that there is a fundamental distinction between a traffic trunk and the path, and indeed the LSP, through which it traverses. An LSP is a specification of the label switched path through which the traffic traverses. In practice, the terms LSP and traffic trunk are often used synonymously.

MPLS - Explicit routing.

One of the current IGP (Internal Gateway Protocol) routing problems is the lack of ability to map traffic trunks into network resources maximizing bandwidth utilization. Another problem is the lack of a mechanism for classifying different classes of service. MPLS, due to explicit routing, avoids these two drawbacks.

Maximizing network resource utilization implies avoiding congestion problems. The next figure shows an example where the use of explicit routing could avoid a classical congestion problem caused by the use of destination-based routing.

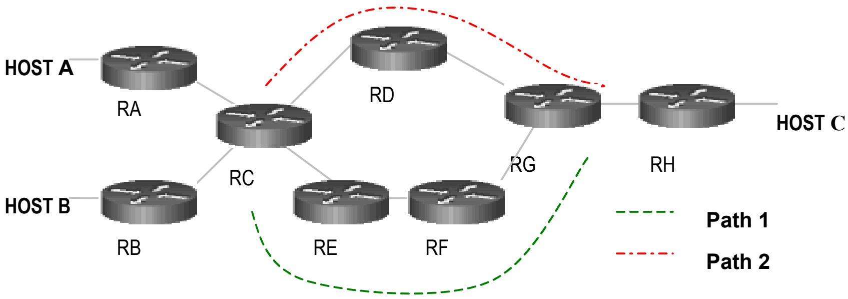


Figure 7. Explicit routing.

Host A, and Host B try to send to Host C. If a short-path routing is used, generated traffic from A and B must be routed by route 1 (path 1). A congestion problem could be detected on node RD, but a short-path routing algorithm could not avoid this problem of splitting traffic from Host A to

Host C and traffic from Host B to Host C. The problems arise from the fact that node C is a destination-based router, so it could not have the ability to detect different traffic classes. With MPLS two explicit routes (two LSPs) could be created between RA-RC-RD-RG-RH (LSP1) and RB-RC-RE-RF-RG-RH (LSP2) solving congestion RD problems.

Constraint-based Routing + MPLS.

Constraint-Based routing (CBR) is the name used to describe QoS routing. A CBR could use several points as input, such as the attributes associated with traffic trunks, the attributes associated with resources, and topology state information. Based on this information, a constraint-based routing process on each node automatically computes the explicit routes for each traffic trunk originating from the node. In this case, an explicit route for each traffic trunk is a specification of a label switched path that satisfies the demand requirements expressed in the trunk's attributes which are subject to constraints imposed by resource availability, administrative policy, and other topology state information.

A constraint-based routing framework can reduce the level of manual configuration and intervention required for updating Traffic Engineering policies. In practice, the Traffic Engineer, an operator, or even an automate-process will specify the endpoints of a traffic trunk and assign a set of attributes to the trunk which encapsulate the performance expectations and behavioral characteristics of the trunk. The constraint-based routing framework is then expected to find a feasible path to satisfy the expectations. If necessary, the Traffic Engineer or a traffic engineering support system can then use administratively configured explicit routes to perform fine-grained optimization.

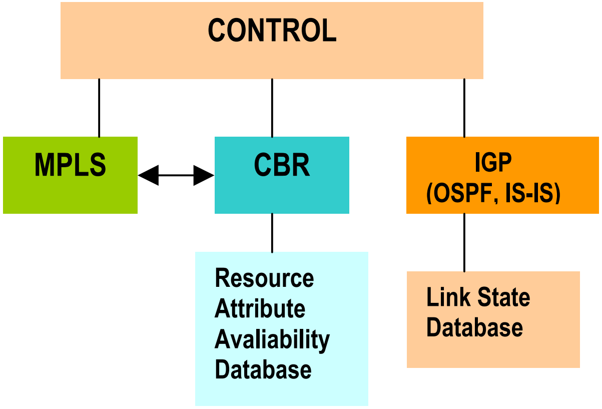


Figure 8. MPLS-CBR

Adding CBR capabilities to MPLS routers (LSR) could be done in two different ways. Firstly it could be done by extending actual IGPs (OSPF o IS-IS) to support CBR capabilities, or secondly by adding CBR as another component that can co-exist with current IGPs.

Loop Detection/Prevention: Another aspect to take into consideration is the loop avoiding system. These situations could occur when a node falls or when the routing protocol is not accurate enough. MPLS propose several mechanisms to prevent loop situations. Some of them are buffer allocation, non-TTL-segments, path-vectors/diffusion algorithm, and colored threads [OHBA-99], [OHBA-01].

1.1.5.3.- IP Virtual Private Networks Services.

Actual solutions to create virtual private networks (VPN) are divided into either: connection orientated protocols (FR or ATM) or TCP/IP mechanisms (tunneling).

With ATM and FR permanent virtual circuits over each VPN nodes are created, causing scalability and management problems (as in the section 1.1.5.1). Tunneling techniques are applied when TCP/IP is used. Those could be conducted at 2 or 3 network level. At 3 level a packet encrypting is used (the most common standard is IPSEC). This means that QoS requirements could not be applied, because IP headers could not be seen to detect QoS packet requirements. At 2 level the encryption is applied over packet frames, allowing QoS application.

The MPLS-VPN (see [ROSE-99] or [DAVI-00]) model, provides all of the advantages of a PVCs model. With this model private network customers could have their own routes and direction planning, avoiding scalability and management problems. Several proposals to develop a MPLS-VPN framework are introduced in the literature (e.g.. [MUTH-00] or [JAMI-98]).

In an MPLS-VPN, LER (label Edge Routers) are called PE (Provider Edge Routers) and LSR (Label Switch Routers) are called P (Provider Routers). Customer nodes are called CE (Customer Edge routers). To exchange routing information between each VPN node an extended BGP (Border Gateway Protocol) for MPLS is used. Figure 9 shows an example of an MPLS Virtual Private Network. In this figure two VPNs (VPN a and VPN b) with their corresponding components are defined in an MPLS scenario.

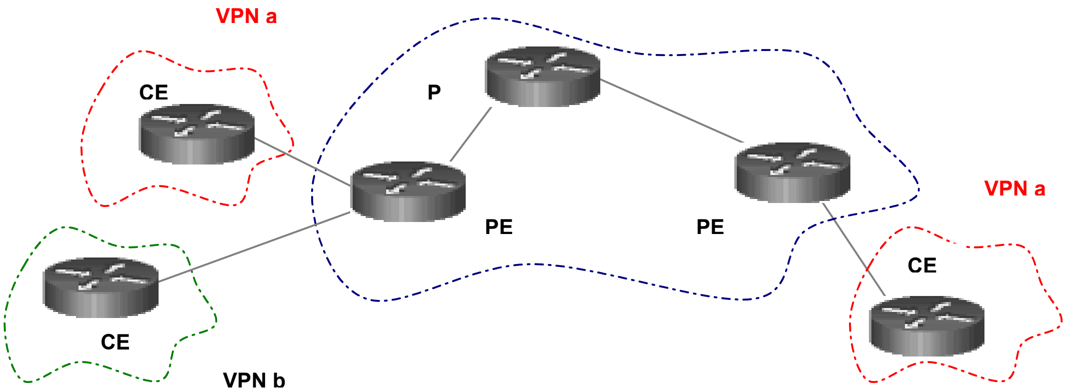


Figure 9. VPN-MPLS Architecture

1.2- MPLS Fault management

1.2.1.- Introduction

MPLS can be used to support advanced survivability requirements and enhance the reliability of IP networks. Differing from classical IP networks, MPLS networks establish label switched paths (LSPs), similar to VP/VC-ATM. This allows MPLS networks to pre-establish protection LSPs, backups for the working LSPs, and achieve better protection switching times than IP networks.

1.2.2.- MPLS protection architecture

The usual method to develop an MPLS protected domain involves a working path and a recovery path (backup path). Not always a backup LSP is created at ingress-node and finalized at egress-node. A backup LSP could be implemented at a LSP segment. In this case the node where the backup is originated, is called a PSL (Path switch LSR) and where the backup ends is called PML (Path Merge LSR).

Components in an MPLS "fault management" mechanism:

Taken in consideration the components described in [SHAR-00], MPLS components are basically the next ones:

1.- A method for selecting the working and the protection paths.
2.- A method for bandwidth reservation for the working and the protection paths.
3.- A method for signaling the setup of the working and protection paths.
4.- A fault detection mechanism to detect faults along a path.
5.- A fault notification mechanism, to convey information about the occurrence of a fault to a network entity responsible for reacting to the fault and taking appropriate corrective action.
6.- A switchover mechanism to move traffic over from the working path to the protection path.
7.- A repair detection mechanism, to detect that a fault along a path has been repaired.
8.- An (optional) switchback or restoration mechanism, for switching traffic back to the original working path, once it is discovered that the fault has been corrected or has been repaired.

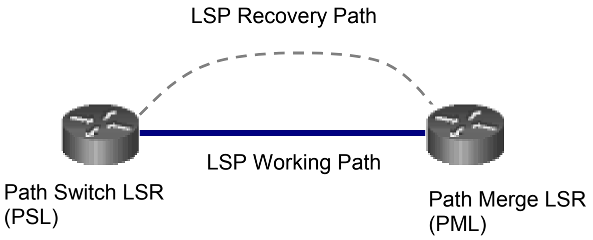


Figure 10. MPLS protected domain

Figure 10 shows a simple MPLS protected domain. This scenario is formed with a Working Path (or a segment of the WP), which is the protected segment and the Backup Path (or the Recovery Path) where the traffic is switched once a failure is detected. The PSL and PML components are two Label Switch routers LSR (LSR) with the protection function.. The next section explains, in more detail, all the components involved in the MPLS fault control.

1.2.3.- Terminology

The next section introduces many definitions to facilitate understanding of MPLS fault management components. All of these definitions are extracted from [MAKA-99]:

MPLS Protection Domain: The set of LSRs over which a working path and its corresponding protection path are routed. The protection domain is denoted as: (working path, protection path).

Fault and recovery signals/messages.

Failure Indication Signal (FIS) : A signal that indicates that a failure has been detected at a peer LSR. It consists of a sequence of failure indication packets transmitted by a downstream LSR to an upstream LSR. It is relayed by each intermediate LSR to its upstream neighbor, until it reaches an LSR that is setup to perform a protection switch.

Failure Recovery Signal (FRS) : A signal that indicates that a failure along the path of an LSP has been repaired. It consists of a sequence of recovery indication packets that are transmitted by a downstream LSR to its upstream LSR. Again, like the failure indication signal, it is relayed by each intermediate LSR to its upstream neighbor, until is reaches the LSR that performed the original protection switch.

Liveness Message (LM) : A message exchanged periodically between two adjacent LSRs that serves as a link probing mechanism. It provides an integrity check of the forward and the backward directions of the link between the two LSRs as well as a check of neighbor aliveness.

Link Failure (LF) : A link failure is defined as the failure of the link probing mechanism, and is indicative of the failure of either the underlying physical link between adjacent LSRs or a neighbor LSR itself. (In the case of a bi-directional link implemented as two unidirectional links, it could mean that either one or both unidirectional links are damaged.)

Loss of Signal (LOS): A lower layer impairment that occurs when a signal is not detected at an interface. This may be communicated to the MPLS layer by the lower layer.

Loss of Packet (LOP) : An MPLS layer impairment that is local to the LSR and consists of excessive discarding of packets at an interface, either due to label mismatch or due to TTL errors.

MPLS protection components.

Working or Active LSP : An LSP established to carry traffic from a source LSR to a destination LSR under normal conditions, that is, in the absence of failures. In other words, a working LSP is an LSP that contains streams that require protection.

Working or Active Path : The portion of a working LSP that requires protection. (A working path can be a segment of an LSP or a complete LSP) The working path is denoted by the sequence of LSRs that it traverses.

Protection Switch LSR (PSL) : An LSR that is the origin of both the working path and its corresponding protection path. Upon learning of a failure, either via the FIS or via its own detection mechanism, the protection switch LSR switches protected traffic from the working path to the corresponding backup path.

Protection Merge LSR (PML) : An LSR that terminates both a working path and its corresponding protection path, and either merges their traffic into a single outgoing LSP, or, if it is itself the destination, passes the traffic on to the higher layer protocols.

Intermediate LSR : An LSR on the working or protection path that is neither a PSL nor a PML.

MPLS Traffic Group (MTG) : A logical bundling of multiple, working LSPs, each of which is routed identically between a PSL and a PML. Thus, each LSP in a traffic group shares the same redundant routing between the PSL and the PML.

Protected MPLS Traffic Group (PMTG) : An MPLS traffic group that requires protection.

Protected MPLS Traffic Portion (PMTP) : The portion of the traffic on an individual LSP that requires protection. A single LSP may carry different classes of traffic, with different protection requirements. The protected portion of this traffic may be identified by its class, as for example, via the EXP bits in the MPLS shim header or via the priority bit in the ATM header.

Protection or Backup LSP (or Protection or Backup Path) : An LSP established to carry the traffic of a working path (or paths) following a failure on the working path (or on one of the working paths, if more than one exists) and a subsequent protection switch by the PSL. A protection LSP may protect either a segment of a working LSP (or a segment of a PMTG) or an entire working LSP (or PMTG). A protection path is denoted by the sequence of LSRs that it traverses.

Protection modes

Revertive : A switching option in which streams are automatically switched back from the protection path to the working path upon the restoration of the working path to a fault-free condition.

Non-revertive : A switching option in which streams are not automatically switched back from a protection path to its corresponding working path upon the restoration of the working path to a fault-free condition.

1.2.4.- Protection types

Protection types for MPLS networks can be categorized as link protection, node protection, path protection, and segment protection.

Link Protection: The objective for link protection is to protect an LSP from a given link failure. Under link protection, the path of the protect or backup LSP (the secondary LSP) is disjointed from the path of the working or operational LSP at the particular link over which protection is required. When the protected link fails, traffic on the working LSP is switched over to protect the LSP at the head-end of the failed link. This is a local repair method that can be fast. It might be more appropriate in situations where some network elements along a given path are less reliable than others.

Node Protection: The objective of LSP node protection is to protect an LSP from a given node failure. Under node protection, the path of the protect LSP is disjointed from the path of the working LSP at the particular node to be protected. The secondary path is also disjointed from the primary path at all links associated with the node to be protected. When the node fails, traffic on the working LSP is switched over to the protect the LSP at the upstream LSR directly connected to the failed node.

Path Protection: The goal of LSP path protection is to protect an LSP from failure at any point along its routed path. Under path protection, the path of the protect LSP is completely disjointed from the path of the working LSP. The advantage of path protection is that the backup LSP protects the working LSP from all possible link and node failures along the path, except for failures that might occur at the ingress and egress LSRs, or for correlated failures that might impact both working and backup paths simultaneously. Additionally, because the path selection is end-to-end, path protection might be more efficient in terms of resource usage than link or node protection. However, path protection may be slower than link and node protection in general.

Segment Protection: An MPLS domain may be partitioned into multiple protection domains whereby a failure in a protection domain is rectified within that domain. In cases where an LSP traverses multiple protection domains, a protection mechanism within a domain only needs to protect the segment of the LSP that lies within the domain. Segment protection will generally be faster than path protection because recovery generally occurs closer to the fault.

1.2.5.- m:n protection model

“m:n protection model” where m is the number of protect LSPs used to protect n working LSPs is one way to classify MPLS restoration models. Feasible protection models could be:

1:1: one working LSP is protected/restored by one protect LSP.

n:1: one working LSP is protected/restored by n protect LSPs, possibly with configurable load splitting ratio. When more than one protect LSP is used, it may be desirable to share the traffic across the protect LSPs when the working LSP fails to satisfy the bandwidth requirement of the traffic trunk associated with the working LSP. This may be especially useful when it is not feasible to find one path that can satisfy the bandwidth requirement of the primary LSP.

1:n: one protection LSP is used to protect/restore n working LSPs.

1+1: traffic is sent concurrently on both the working LSP and the protect LSP. In this case, the egress LSR selects one of the two LSPs based on a local traffic integrity decision process, which compares the traffic received from both the working and the protect LSP and identifies discrepancies. It is unlikely that this option would be used extensively in IP networks due to its resource utilization inefficiency. However, if bandwidth becomes plentiful and cheap, then this option might become quite viable and attractive in IP networks.

Recovery Paths types with QoS requirements

Equivalent Recovery Path : Means that recovery path preserve QoS Working Path requirements.

Limited Recovery Path : Does not preserve QoS requirements.

Fast MPLS recovery

Usually, Fast recovery methods are associated with fast recovery in terms of time. Is difficult to establish a height of this time. Actually, this height use to be associated with the SONET/SDH recovery times (less than 50 ms). MPLS with pre-established backups promise to obtain similar times [OWEN-00].

1.2.6.- Network Survivability Layer Considerations

While best effort networks were focussed primarily on connectivity, means re-routing fault management systems were enough to provide survivability, actual networks begin to support different classes of services (critical traffic, real-time traffic or high priority traffics), which means that slow re-routing schemes are not enough to achieve reliable fast services. The main drawback of level 3 re-routing algorithms is the amount of time that the algorithms took to converge and restore service. Actual networks need to provide highly reliable services, where the time needed to recover a failure might be of the order of milliseconds. In practice, fault restoration capabilities are implemented in multiple protocol layers, such as automatic protection switching in the physical transmission layer, self-healing in the ATM virtual path layer, and fast rerouting in MPLS [CHEN-99]. Usually, fault recovery is attempted firstly at the lowest layer, and then escalated to the next layer if recovery was unsuccessful or not possible.

To achieve fault management actual networks provide different schemes at different layers [OWEN-00]. At the bottom of the layered stack (optical networks) ring and mesh topology restoration functionality at the wavelength level, is provided. At the SONET/SDH layer survivability is provided at a link level in ring and mesh architectures. Similar functionality is provided by layer 2 technologies such as ATM (generally with slower mean restoration times). Rerouting is traditionally used at the IP layer to restore service following link and node failures. Rerouting at the IP layer occurs after a period of routing convergence, which may require anything from seconds to minutes to complete. MPLS allows new restoration mechanisms, with better performance than IP re-routing mechanisms. Recently, a common suite of control plane protocols has been proposed for both MPLS and optical transport networks under the acronym Multiprotocol Lambda Switching (MPλS) [AWDU-99d]. This new paradigm of Multiprotocol Lambda Switching will support even more sophisticated mesh restoration capabilities at the optical layer for the emerging IP over WDM network architectures.

Developing a multi-layer survivability scheme involves providing restoration at different time scales (temporal granularity). Bandwidth granularity is another way of classifying protection mechanisms. Bandwidth granularity goes from the wavelength level (optical level) to packet-level (IP and higher layer protocols). Another vision of protection applicability is from the point of view of network services or traffic classes.

General requirements for protection and restoration coordination.

Protection and restoration coordination across layers may not always be feasible, because networks at different layers may belong to different administrative domains. Several points at which to minimize the impact of different layer protection disruption to achieve an efficient and complete protection scheme are according to [OWEN-00]:

- Minimization of function duplication across layers is one way to achieve coordination. Escalation of alarms and other fault indicators from lower to higher layers may also be performed in a coordinated manner. A temporal order of restoration trigger timing at different layers is another way to coordinate multi-layer protection/restoration.

- Spare capacity at higher layers is often regarded as working traffic at lower layers. Placing protection/restoration functions in many layers may increase redundancy and robustness, but it should not result in significant and avoidable inefficiencies in network resource utilization.

- It is generally desirable to have protection and restoration schemes that are bandwidth efficient.

- Failure notification throughout the network should be timely and reliable.

- Alarms and other fault monitoring and reporting capabilities should be provided at appropriate layers.

The next table introduces several main fault management features of each network level introduced in [OWEN-00]:

Protection principles in the network layers	
Optical Layer	• "Fast fault failure detection": the loss of light or carrier signals detection and switching to a backup lightpath (if configured). • Limited at lighpath granularity. • No discrimination between traffic types.
Sonet/SDH Layer	• Limited to ring topologies and may not always include mesh protection. • Cannot distinguish between different priorities of traffic. • Not vision of higher layer failures. • Limited to link failures
ATM Layer	• Node failure detection (F1-F5 its mechanisms, "peer capabilities") • "in band OAM functionality" : fast path error detection. • "Mis-configurations" detection: VPI/VCIs errors.
MPLS Layer	• Node/link failure detection: "Path Continuity Test", "Fast Liveness Message Test" • "Mis-configuration" errors: unlabeled packets, unrecognized labels, TTL (Time to Live) mismatches.
IP Layer	• Re-routing mechanisms (too slow).

Table 1: Fault management features at each network level.

1.3.- MPLS fault management mechanisms.

The usual method to offer protection in MPLS environments is to pre-establish a backup LSP to switch back the traffic when failure occurs. Backup types could be different depending on where they have originated or what types of fail/recovery notification are activated. This section is merely an introduction of different type of LSP backups and their notification methods, most of them have been proposed in different IETF drafts such as [HUAN-00], [KINI-00], [HASK-00], [KRIS-99] or [31]. Their main principles and a review of the pros and cons are introduced in this section.

1.3.1.- Centralized model

In this model, an Ingress Node takes the responsibility to resolve the restoration as the FIS (Fault Indication Signal) arrives. This method needs an alternate disjoint backup path for each active path (working path).

Protection is always activated at the Ingress Node, irrespective of where along the working path a failure occurs. This involves that the failure information has to be propagated all the way back to the source node before a protection switch is activated. If no reverse LSP is created the fault indication can only be activated as a Path Continuity Test.

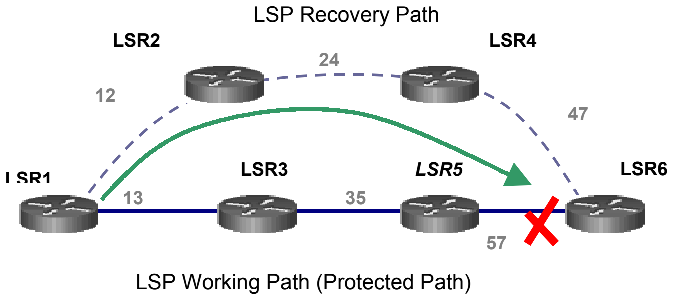


Figure 11 : Centralized model

This method has the advantage of setting up only one backup path per working path, and is a centralized protection method which means that only one LSR has to be provided with PSL functions. On the other side this method has an elevated cost (in terms of time), especially if a Path Continuity Test is used as a fault indication method. If we want to use an RNT as a fault indication method we have to provide a new LSP to reverse the signal back to the Ingress Node.

Figure 11 shows a simple scenario formed by six LSRs where a working path (i.e: LSR1-LSR3-LSR5-LSR6, the solid line) and a LSP recovery path (i.e: LSR1-LSR2-LSR4-LSR6, the dashed line) are pre-established,. In a normal operation, traffic from ingress router LSR1 to egress router LSR6 is carried through the LSP working path. When a link fault is detected, (for instance between LSR5 and LRR6), traffic is then switched to the LSP Recovery Path; the arrow shows this new path.

1.3.2.- LSP segment restoration (local repair)

With this method restoration starts from the point of the failure. It is a local method and is transparent to the Ingress Node. The main advantage is that it offers lower restoration time than the centralized model.

An added difficulty, of the local restoration, arises in that every LSR, where protection is required, has to be provided with switchover function (PSL). A PML should be provided too. Another drawback is the maintenance and creation of multiple LSP backups (one per protected domain). This could report low resource utilization and a high development complexity. On the other hand, this method offers transparency to the Ingress Node and faster restoration time than centralized mechanisms.

Figure 12 illustrates this case, the same working path as in centralized model is used (i.e: LSR1-LSR3-LSR5-LSR6, solid line). The LSP recovery path is now formed by LSR3-LSR4-LSR6 that is shorter than the LSP recovery path in the centralized method. When a link failure occurs, traffic is switched from LPD (LSR5-LSR6) which is a segment of the working path to the LSP recovery Path.

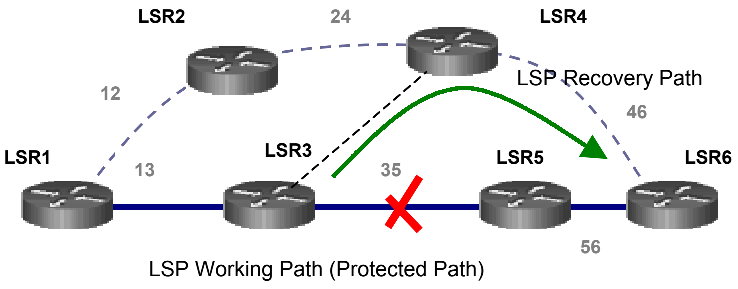


Figure 12 : Local restoration

An intermediate solution could be the establishment of local backup, but only for protection segments where a high degree of reliability is required, supplying only protected path segments.

1.3.3.- Reverse backups

Pre-established alternative paths are essential where packet loss due to an LSP failure is undesirable. Since it may take a significant time for a device on a label switched path to detect a distant link failure, it could continue sending packets along the primary path. As soon as such packets reach a switch that is aware of the failure, the switch to an alternative path away from the failure must immediately reroute packets if loss of data is to be avoided.

The main idea of this method is to reverse traffic at the point of a failure of the protected LSP back to the source switch of the protected path (Ingress Node) via a Reverse Backup LSP.
As soon as a failure along the protected path is detected, the LSR at the ingress of the failed link reroutes incoming traffic. It redirects this traffic into the alternative LSP traversing the path in the reverse direction of the primary LSP.

This method is especially suitable in network scenarios where the traffic streams are very sensitive to packet losses. For example in a voice transmission, delay is one of the main aspects, but if a file is transmitted, packet losses could be critical. If the link segment or the node where the failure occurs is situated far from the ingress node and the transmission rate is very high, the number of packet lost could be very high if a centralized backup is used. Reverse backup utilization allows the recovery of packets as the failure occurs, rescuing lost packets if a centralized method is applied.

Another advantage is that it simplifies the fault indication, since the reverse backup offers, at the same time, a way to transmit the FIS to the Ingress Node and the recovery traffic path. One disadvantage could be poor resource utilization. Two backups per protected domain are needed. Another drawback is the time required to reverse fault indication to the Ingress Node as in the Centralized model. Regardless, a reverse backup can be established in association with the working path, simply by making each LSR along a working path remember its neighbor.

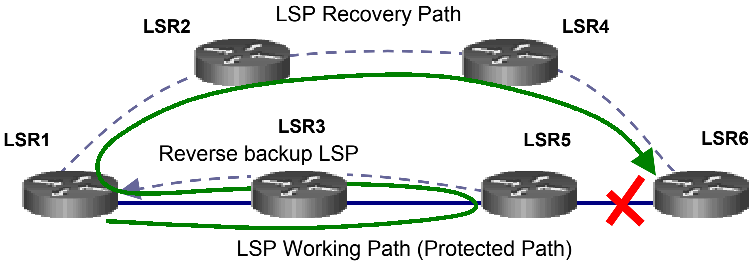


Figure 13 : Reverse backup utilization

Figure 13 shows an example of reverse backup utilization. LSP working and recovery paths are established as in the centralized model, in addition there is a reverse path from LSR5 (LSR5-LSR3-LSR1) which reaches the ingress node. When a link failure is detected in LSP (LSR5-

LSR6), the traffic is switched back to LSR1 (ingress node) through the reverse backup LSP, and then carried through the LSP recovery path as in the centralized model.

1.3.4.- Fault notification

One of the main points in the recovery cycle is the detection and notification of link/nodefailures. Fault detection could be done at different network layers, depending on the type of failure and the type of lower layer protocols. Fault notification, once a failure is detected, could be localized or centralized (see previous section). If a local restoration method is used Fault notification, does not usually have to be done, because actions for recovery are taken at the same node that detected the fault. This is not entirely true, as the local protection could cover a path segment and the failure notification does not necessarily have to be done by the node responsible for the switch over operation. On the other hand , if it is an ingress node, or a node that is not necessarily the one responsible for the fault detection, the fault must be communicated from the point of failure to the ingress node or to the node designed to trigger recovery actions (PSL nodes). Reverse Notification Tree (RNT), a proposal introduced in [HUAN-00], is one proposal to develop notification in a centralized or segment protection environment. This section explains the main principles and goals of this proposal.

Reverse Notification Tree

The reverse notification tree is a point-multipoint tree rooted at the PML (Path Merge LSR) along which the FIS (Fault Indication Signal) or a FRS (Fault Recovery Signal) travels to a PSL (Path Switch LSR). Using a Reverse Notification Path (RNT) method gives the following advantages:

- The RNT can be established in association with the working path, simply by making each LSR along a working path remember its upstream neighbor (or the collection of upstream neighbors whose working paths converge at the LSR and exit as one). No multicast routing is required.
- Only one RNT is required for all the working paths that merge to form the multipoint-to-point forward path. The RNT is rooted at the PML and terminated at the PSLs. All intermediate LSRs on the converged working paths share the same RNT.

Protection Domain: Different backup types could be established to offers MPLS protection. In many cases the network topology could be forced to setup only one type of backup, for example, in Figure. 14, the working path 9-3-4-6-7, can only have protection on the segment 9-10-7. Both centralized and segment protections are taken into account to develop a notification method proposal.

Relationship between protection domains

Multiple LPSs could merge into a single LSP. In this case, it would propagate the failure (and the recovery) notification back to the concerned PSL(s) involved in developing a reverse notification tree. Two scenarios could happen depending on whether the protection domains are independent of each other or not. For example, the protection domain defined by (9-3-4-6-7, 9-10-7) is completely independent of the domain defined by (13-5-15, 13-14-15). Once a failure occurs that failure does not affect the other RNT, so therefore multiple failure detection could be done at the same time.

If protection domains with different RNTs overlap, failures on the working paths of the two domains do not affect one another, due to the fact that each RNT works independently of each other. However, failures on the protection path of one may affect the working path of the other and visa versa. For example, the protection domain defined by (1-2-3-4-6-7, 1-5-7) is not independent of the domain defined by (11-13-5-15, 11-13-14-15) since LSR 5 lies on the protection path of the former domain and on the working path of the latter domain.

When protection domains have the same RNT, different failures along the working paths may affect both paths differently. As shown in Figure 14, for example, working paths 1-2-3-4-5-7 and

9-3-4-6-7 share the same RNT. As a result, for a failure on some segment of the working path, both domains will be affected, resulting in a protection switch in both (for example, the segment 3-4-6-7 in Fig. 14). However, for failures on other segments of the working path, only one domain may be affected (for example, failure on segment 2-3 affects only the first working path 1-2-3-4-6-7, where as failure on the segment 9-3 affects only the second working path 9-3-4-6-7).

Path Protection Operation

The following sections, describe the operation of a path protection mechanism, explaining the various steps involved with reference to Fig. 14.

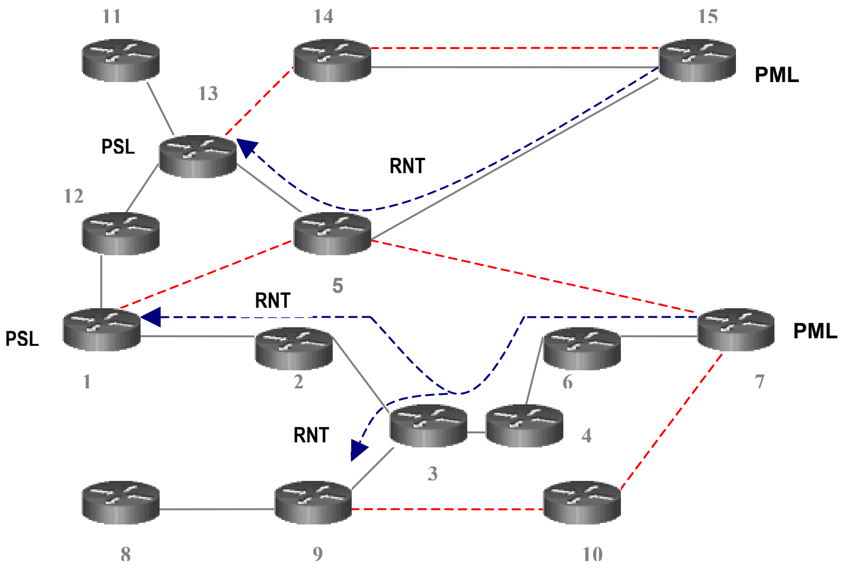


Figure 14: Illustration of MPLS protection configuration.

Different timers and thresholds are defined to develop the proposal. This timer controls several aspects as the maximum duration of each operation (protection switch, restoration switch…) or intervals between different packets, etc (for more details see [HUAN-00]). The next section explains how to create an RNT and a fault detection/notification is done (a complete protection cycle is explained in [HUAN-00]).

Creating the protected paths and the RNT

Protection configuration consists of two aspects: establishing the protection path and creating the reverse notification tree. The establishment of the protection path involves assigning different function to the path components. These functions are, more or less, the usual functions of a PSL, PML and the intermediate LSRs.

The RNT is used for propagating the FIS and the FRS, and can be created very easily by a simple extension to the LSP setup process. During the establishment of the working path, the signaling message carries with it the address of the upstream node that sent it. Each LSR along the path simply remembers the identity of its immediate prior upstream neighbor on each incoming link. The node then creates an inverse cross-connect table that, (for each protected outgoing LSP) maintains a list of the incoming LSPs that have merged into that outgoing LSP, together with the identity of the upstream node that each incoming LSP comes from. Upon receiving an FIS, an LSR extracts the labels contained in it (which are the labels of the protected LSPs that use the outgoing link that the FIS was received on) consults its inverse cross-connect table to determine the identity of the upstream nodes that the protected LSPs come from, and creates and transmits an FIS to each of them.

Basically, the main associated functions to the protected path components are:

PSL: The PSL must be able to correlate the RNT with the working and protection paths. To this end, it maintains a table with a list of working LSPs protected by an RNT, and the identity of the protection LSPs that each working path is to be switched to in the event of a failure on the working path. It need not maintain an inverse cross-connect table (for the LSPs and working paths for which it is the PSL).

PML: The PML is the root of the RNT, and has to associate each of its upstream nodes with a working path and RNT. It need not maintain an inverse cross-connect table (for the LSPs and working paths for which it is a PML).

Intermediate LSR: An intermediate LSR has to only remember all of its upstream neighbors and associate them with the appropriate working paths and RNTs. It has to maintain an inverse cross-connect table.

Failure notification

Each LSR must be able to detect certain types of failures and propagate an FIS message towards the PSL. A complete analysis of how each fault type has to be managed by each protection component is introduced can be seen in [HUAN-00]. They consider the failures: unidirectional link failure, bi-directional (or complete) link failure, and node failure.

The notification method acts as a failure is detected. For instance if a failure (in the link 23) is detected by the LSR 3 an FIS is sent to LSR 2. The FIS will contain the incoming label of those LSPs on link 23. Upon receiving the FIS message, LSR 2 will consult its inverse-cross-connect table and generate an FIS message for LSR 1, which on receiving the first FIS packet will perform the switch over action.

Basically, the main associated functions to the protected path components are:

PSL: Detect FIS packets.

PML: Generate FIS packets and transmit them over the RNT.

Intermediate LSR: Must be able to generate FIS packets (in response to a detected failure or a received FIS packet). It must transmit these to all its affected upstream neighbors as per its inverse-cross-connect table.

1.3.5.- Shared backups

Using a disjoint backup path for a working LSP is the common way to provide reliability. However this requires at least twice the amount of network resources. Backup paths could be shared between different working paths in a way that single link/node failure recovery is guaranteed providing a good network resource utilization. A proposal to route shared backup paths is introduced in [KINI-00]. This proposal routes these backups using only aggregated network usage information (this is extended in [KAR-00] as and is explained in the next section). In this section several examples of shared backup utilization are reviewed. How to route and setup these shared backups (using an on-line routing algorithm and a signaling method) are explained in more detail in the next section.

Shared backup examples

Several examples of shared backup applications to recover different network failures are introduced in this section.

Single link/node failure recovery.

Figure 15 shows a simple case of sharing backup paths to recover single link failure. Say each link is of unit bandwidth and each LSP request is also of unit bandwidth. L1 and L2 are two working paths. L1b is the backup for L1 and L2b is the backup for L2. L1b and L2b can be placed on the same link by sharing the bandwidth. Clearly, if either one of L1 or L2 fail the system can recover using the shared backup.

Figure 16 shows that a simple case of sharing backup paths to recover single node failure can be recovered. L1 is a working path along the label switch routers E-F-G. The corresponding backup L1b is along the path E-C-D-G. Similarly L2 is an active path along A-B. L2b is the corresponding backup path along the Label switch routers A-C-D-B. Clearly, if max-bandwidth (L1,L2) is allocated on link C-D for L1b and L2b together, the system can ensure single node failure recovery.

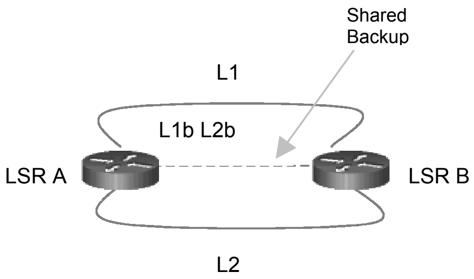


Figure 15 : Sharing backup links with link failure recovery

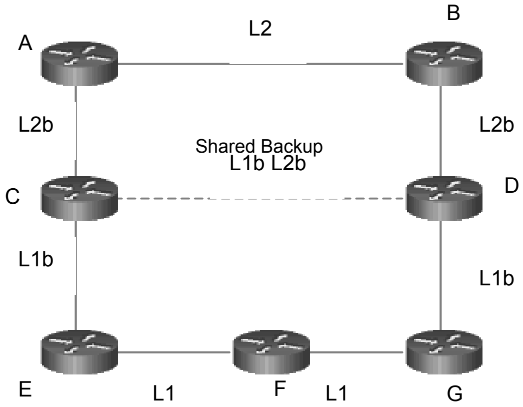


Figure 16 : Sharing backup links with node failure recovery

Shared backups with local restoration

Local restoration can be achieved by providing intermediate nodes with a backup path (see previous sections). Figure 17 illustrates an example of local restoration for single link failure recovery. Sharing of backup paths can be done in this case to achieve single link failure recovery. Sharing of links between segments of the backup paths, along the label switch routers A-D-B and B-E-C, could be done to achieve better resource utilization. Other examples of shared backup utilization in the case of single node failure restoration (with local backups) can be found in [KINI-00].

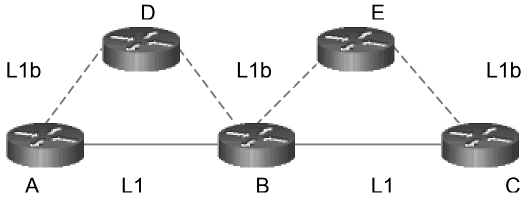


Figure 17 : Local restoration of link failure

A simple algorithm for calculated shared backup path

Routing a backup and a working path guaranteeing QoS requirements to achieve a good network performance is an important aspect. In [KINI-00] an algorithm to compute the bandwidth allocation of both working and backup paths is introduced.

Terminology : For example for link (i,j)

1.- the cumulative bandwidth allocated for active paths is F(i,j)
2.- the cumulative bandwidth allocated for backup paths is G(i,j)
3.- the residual bandwidth free for allocation is R(i,j)

For a request of bandwidth b the active path is calculated as the shortest path on the topology of links that have R(i,j) > b. Let M be the max of the F values along the active path. The backup path is calculated as follows. The cost of a link (u,v) is now taken as

1.- 0 if { M+b < G(u,v) } else
2.- b if { G(u,v) <= M and b <= R(u,v) } else
3.- M+b - G(u,v) if { M <= G(u,v) and M+b <= G(u,v)+R(u,v) } else
4.- infinity in all other cases

The backup path is calculated as the shortest path on the topology with the cost of links calculated as above. The information needed to develop this algorithm has to be proportioned by the routing protocol. Aggregate information about a link that has to be conveyed by a link state routing protocol should consist of

1.- The total bandwidth used on the link for active LSPs
2.- Total bandwidth used on the link for backup LSPs
3.- Total available bandwidth on the link

This algorithm is only a proposal to compute a QoS path (in this case taking in consideration the restoration case). The next section (1.4) describes with more detail how LSPs, not only backup LSPs, can be route guaranteeing certain QoS parameters.

1.4.- MPLS QoS on-line routing.

1.4.1.- Routing algorithms

Routing algorithms attempt to find a feasible path. These algorithms could be divided depending on what type of routing information is used to compute path routes and when this computation is applied. Firstly taking into account that classification routing algorithms could be statics or dynamics Then static algorithms only use static network information, while dynamic algorithms use link load information what is actualized periodically. Secondly, routing algorithms could be, depending when paths are computed, on-line routing (or on-demand routing) and off-line routing (or pre-computed routing). With on-line routing algorithms path requests are attended to one by one, while off-line routing does not allow new path route computation (because there are pre-computed).

1.4.2.- QoS routing. Principles and Previous work

The main goal of a routing algorithm is to find a feasible path (a path with enough bandwidth) that achieves efficient resource utilization. To optimize network performance QoS routing algorithms use two techniques. The firsts one is to pick the minimum hop count path in order to reduce the resource consumption, or alternatively, to load balance the network the least loaded path is selected. This optimization of the network utilization: reducing resource consumption and balancing the network load, is not easy to be achieved using only a unique routing algorithm since these two objectives use to be opposites. That means a path with the least number of hops does not necessarily have to be the path with the best resource consumption. This is the reason why developing a suitable QoS algorithm involves taking into account more than one aspect. A good way to develop a suitable QoS routing algorithm, with the objectives of load balance and resource consumption in mind, is to develop new routing criterias or to mix several QoS criteria. These QoS criterias could be, apart from minimum hop count, the maximum residual bandwidth, the minimum path cost based on the link utilization, etc. In the literature several proposals of QoS routing taking into consideration these criterias or mixing many of them are developed and experimented with [GUER-97],[MA-97].

A usual routing method is to use a min-hop algorithm (MHA). This algorithm only chooses the feasible path with the least number of hop (links) as a unique routing criteria. In [GUER-97] a widest-shortest path (**WSP**) algorithm based on the Bellman-Ford algorithm is proposed. They mix the two criterias. The first one is to pick the path with the minimum hop count amongst all feasible paths. If more than one path is chosen the one with the maximum reservable bandwidth (MRB) is selected. The MRB on a path is the minimum of the reservable bandwidth of all links on the path. Another routing proposal is exactly the opposite of the WSP, that means the first criteria is the path with the minimum bandwidth and if more than one is feasible the one path with the minimum hop count is then selected. This algorithm is called the shortest-widest path (**SWP**). If in these last proposals one (WSP) gives the highest priority to the resource utilization and the other (SWP) one gives its priority to balancing the network load other proposals define a cost function and applies a shortest-path algorithm based on this cost. Last algorithms present several drawbacks to selecting a path with a longer number of hops (in the case of WSP) or a path with a critical bandwidth allocation, that could become a congested point. To avoid this, other proposals impose constraints, which acts to relax these drawbacks. In Dynamic-alternative path (**DAP**) [MA-97], a hop count restriction to avoid selecting a path with n units superior to the number of hops computed by MHA is used. Is basically the WSP with a hop limit.

Several proposals that make use of MPLS network capabilities to develop new path selection algorithms with QoS guarantees are proposed in the recent literature ([KODI-00], [KAR-00], or [SURI-00]). In these proposals, as a difference with last QoS routing algorithms, the use of ingress-egress nodes knowledge is the common denominator.

1.4.3.- MPLS QoS on-line routing algorithms

MPLS due to its capabilities facilitate the implementation of QoS parameters to route new paths (LSPs). In this section a review of several MPLS QoS on-line routing proposals is introduced. Their advantages and disadvantages are pointed out.

Dynamic Routing of bandwidth guaranteed tunnels with restoration.

This is one of the first proposals [KODI-00] that take into consideration MPLS aspects to design a routing proposal. They develop an on-line routing algorithm of bandwidth guaranteed LSPs to route backup and working paths as a request arrive. In their algorithm if sufficient bandwidth is not available to setup either the active or the backup path then the request is rejected. They consider only the case of protection against simple link/node failures. The case of multiple backup establishment is not considered, quite the contrary the possibility of sharing backups is one of the main points of this paper.

Different routing methods, based on the information available to path computing, are explained. These methods compute, basically, an integer linear programming problem. An algorithm with only aggregated link bandwidth usage information (called dynamic routing with partial-information **DR-PI**) is principally proposed as a good solution in terms of compute cost and performance.

The main goal of this proposal is to develop an on-line routing algorithm to minimize bandwidth usage. The difference to other proposals with this method does not have in their priorities the minimizing of the request rejection. Nevertheless an study of the blocking rate between the proposed algorithms take as a result, similar to [MA-97] experiments, that if the routing algorithm has better knowledge of the actual network parameters, less rejected requests are computed. The main conclusion of this proposal is that an algorithm with only aggregated link bandwidth usage information performs as well as algorithms with more complete information, in terms of bandwidth allocation.

Main drawback of this proposal is that the request rejection counting or the request of multiple backups (or simple an LSP request) is not taken into account. This drawback is improved in the next proposal.

Minimum Interface Routing Algorithm

In the “Minimum Interface Routing Algorithm” (**MIRA**) [KAR-00] and [AUKI-00], another proposal that takes into consideration aspects of MPLS architecture to design a on-line routing scheme. In this case, ingress and egress nodes are taken into account, is introduced. Kodialam and Lakshman introduce the concept of interference, and develop a multiple max-flow computation to determine the path of least interference.

Interference

The main idea is to establish paths that do not interfere “too much” with future LSP setup requests, considering pre-established values ingress-egress pairs. Figure 18 shows an example of this “interference” effect. Consider the maximum flow (maxflow) value 1 between a given ingress-egress pair (S1, D1). Note that maxflow value 1 decreases whenever a bandwidth demand is routed between S1 and D1. The value of 1 can also decrease when an LSP is routed between some other ingress-egress pair. They define the amount of interference on a particular ingress-egress pair, say (S1, D1), due to routing an LSP between some other ingress-egress pair as the decrease in the value of 1.

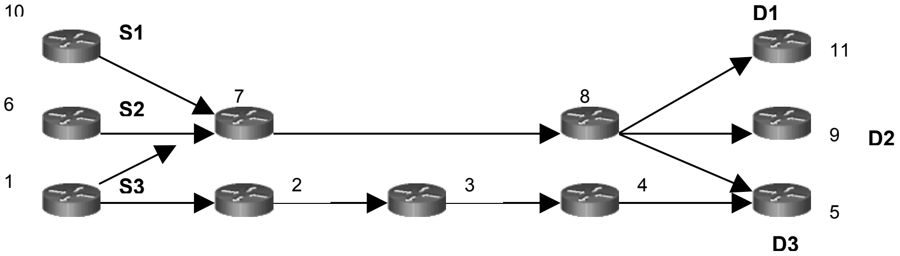


Figure 18 : Minimum Interference Paths

Existing LSP1 (S1,D1) and LSP2 (S2,D2) and LSP3 is required between S3 and D3. If MHA (Minimum Hop Algorithm) is used, the route between (S3,D3) will be 1-7-8-5. This route produces a blocking path between S2 and D2 as well as S1 and D1. In this example it is better to choose route 1-2-3-4-5 even though the path is longer.

Minimum Interference Paths

The minimum interference path for an LSP between, say, (S1, D1), is the explicit route which maximizes the minimum maxflow between all other ingress-egress pairs. In another words, this can be thought of as a choice of path between (S1, D1) which maximizes the minimum residual capacity between every other ingress-egress pair.

The objective might be to choose a path that maximizes a weighted sum of the maxflows between every other ingress-egress pair. This algorithm not only makes capacity available for the possible arrival of future demands, but also makes capacity available for rerouting LSPs in case of link failures.

Critical Links

Critical links are links with the property that whenever an LSP is routed over them, the maxflow value of one or more ingress-egress pairs decrease. This is the criteria to create a weighted graph.

Path Selection by Shortest Path Computation

They use Dijkstra or Bellman-Ford algorithms for computing actual explicit route. They do this by generating a weighted graph where the critical links have weights that are an increasing function of their criticality. The increasing weight function is picked to defer loading of critical links whenever possible. The actual explicit route is calculated using a shortest path computation as in other routing schemes.

The algorithm has an input graph G(N,L) and a set B of all residual link capacities. An ingress node a and an egress node b between which a flow of D units have to be routed. And generate an output route between a and b having a capacity of D units of bandwidth.

MIRA drawbacks

An experimental analysis of MIRA [OWEN-00] points out that in a set of network scenarios MIRA does not work as expected. Two main drawbacks are highlighted in the following:

MIRA focuses exclusively on the interference effect on single ingress-egress pairs. For example figure 19 illustrate this effect. In [OWEN-00] this network is called “The concentrator topology”.

One node C acts as a concentrator for n ingress nodes S1..Sn. Node C is connected to a high capacity link of capacity n+1, whose endpoint is an egress node D. A high bandwidth ingress node S0 is also connected to the concentrator, through a n capacity link. S0 is also connected to D via an alternative 3-hop path, of capacity n.

In this example the MIRA checks the LSP requests one by one. The first request (S0,D) has two possible paths (S0,C,D): 2-hops and (S0,E,F,D): 3-hops. The first one is not considered so critical because it is not considered a minimum cut for any individual ingress-egress pair this permits a residual bandwidth 1, enough for any individual request. Therefore, MIRA chooses the path (S0, C, D) which is an incorrect path in this scenario. An optimal algorithm would route the (S0,D) request along the top on the alternative path (S0,E,F,D), and it would use the (C,D) link to route the n 1-unit request from Si to D. More examples of this drawback are shown in [4]. Other examples of this effect are shown in [SURI-00].

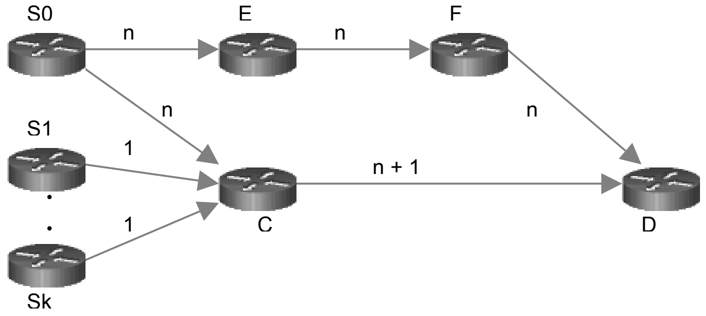


Figure 19: The concentrator topology

Another drawback is that MIRA is computationally very expensive. MIRA performs hundreds of maximum flow computations, each of which is several orders of magnitude more expensive than shortest path computations.

Profile-Based Routing: A new Framework for MPLS Traffic Engineering.

Suri, Waldvogel and Warkhede introduce, in [SURI-00], the idea of using a "traffic profile" of the network, obtained by measurements or service level agreements (SLAs), as a predictor of the future traffic distribution. The objective is that algorithm could anticipate a flow's blocking effect on groups of ingress-egress pairs (MIRA only considers one ingress-egress pair at a time).

The ability, of MPLS networks, to specify explicit paths for any flow gives an important tool to engineer how traffic is routed, and thereby improve the network utilization, by minimizing the number of requests that are rejected when the network becomes overloaded. A traffic profile can be as simple as an average bandwidth requirement over a certain time period.

The Profile-Based Routing (PBR) uses quasi-static information in a preprocessing step (one multi-commodity flow computation), to determine certain bandwidth allocations on the links of the network. The on-line phase of the routing algorithm then routes LSP requests using a "shortest path" (SPF) like algorithm but with additional information given by the preprocessing phase. The multi-commodity-preprocessing phase allows the on-line algorithm to exercise admission control by rejecting some requests because of their blocking effects in the network.

The multi-commodity flow formulation permits a cost function, which they minimize to achieve optimal routing. In order to minimize the number of rejected requests, they use the simple "linear cost function". A variety of non-linear cost functions can be used to handle features such as minimum guaranteed bandwidth or fairness across multiple flows.

One drawback, of this proposal, is the no explicit recovery treatment. As in the case of MIRA only ingress-egress nodes are considered. In MIRA only the case of a centralized backup establishment (one backup along a path formed of a source ingress-node and a destiny egress-node) is considered, no local or reverse backups are considered. In PBR any type of backup establishment is considered.

The next table introduces a taxonomy of the methods reviewed in the last section:

	Algorithm	Refs.	Main objective	Routing Information	Route computation	Drawbacks
QoS routing algorithms	WSP (Widest-Shortest Path)	[GUER-97] [MA-97]	Gives highest priority to resource utilization.	Maximal reservable bandwidth (MRB).	MHA over feasible paths first and the path with the maximum-reservable bandwidth.	Select a path with a longer number of hops (only in the case of the WSP). No limit is established. Select a path that could become a congested point (no request rejection aspect is considered). No recovery treatments are considered.
	SWP (Shortest-Widest Path)	[MA-97]	Gives highest priority to balancing the network load.		The path with the MRB first and the MHA path over the MRB results	
	DAP (Dynamic Alternative Path)	[MA-97]	Improve WSP limiting the path hop/link number.		A WSP with a hop count restriction	
MPLS on-line routing algorithms	DR-PI (Dynamic Routing with Partial-Information)	[KODI-00]	Optimize the bandwidth usage.	Ingress-Egress Nodes and the aggregated link bandwidth usage.	An integer linear programming problem	The numbers of rejected request are not taken in consideration. No local/segment backups are considered.
	MIRA (Minimum Interference Routing Algorithm)	[KODI-00] [AUKI-00]	Optimize the bandwidth usage and minimizing the number of rejected request.	Ingress-Egress nodes and link bandwidth usage.	The concept of the interference generates a weighted graph with the critical links (as a cost) and a SPF algorithm picks the path.	Cannot detect critical links in topologies with clusters of nodes Computationally expensive. No local/segment backups are considered.
	PBR (Profile-Based Routing)	[SURI-00]	Optimize the bandwidth usage and minimizing the number of rejected request	Ingress-Egress nodes. Current residual capacity. Traffic class (service type).	A pre-processing step (multi-commodity flow computation) to determine certain BW allocation and an on-line phase using a SPF algorithm.	No explicit recovery treatments are considered.

Table 2 : QoS routing algorithms.

1.4.4.- Simulation scenario

One important aspect to develop a performance analysis of any mechanism is to design a simulation scenario, which is the same for each test. In [KODI-00] and [KAR-00] (the proposals explined in the last section) a network scenario (see fig. 20) is defined. Afterwards in [SURI-00] this scenario is referenced as the KL-graph, and their experimentation is applied to it.

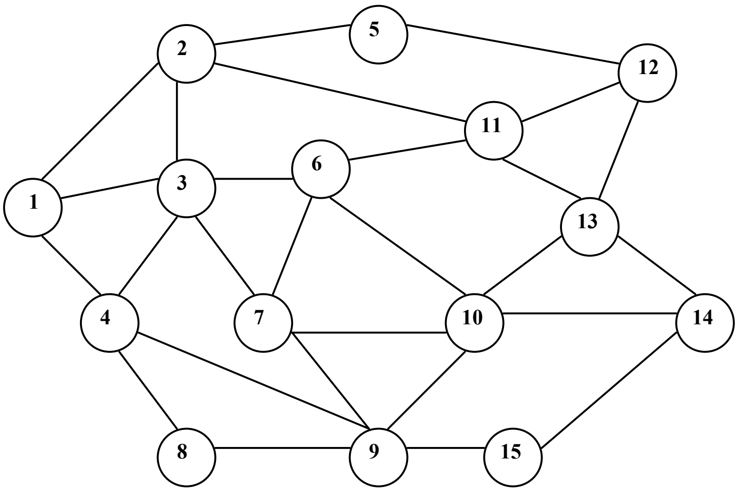


Figure 20. Node test network (KL-graf)

2 Chapter

Problem specification and thesis proposal

2.1.- Problem specification

MPLS allows packet encapsulation at network ingress points (ingress nodes) labeling packets and routing these packets along LSP (Label Switch Paths). These LSPs could be seen as traffic trunks, carrying aggregated streams, classified into FECs (Forwarding Equivalence Classes). These classification/aggregation streams added with other MPLS capabilities (especially with explicit routing, which defines which nodes have to be part of an LSP), allows MPLS to be a powerful tools to provide TE (Traffic Engineering) actual networks.

In a first phase a network could be engineered, but network characteristics change. Generally, network resources could change due to new resource requests or topology changes (such as node or link failures). In these cases a new dynamic traffic-engineering plane has to be triggered. One important part of designing a QoS network is the reliability of the network. This reliability could be provided with different fault management mechanisms. These mechanisms are applied at different network levels. MPLS provides a fast restoration method to recover failures. MPLS fault restoration mechanisms usually use backup LSPs establishment. With these backups, traffic could always be redirected when failure occurs.

MPLS allows failure detection and failure recovery actuation, fast and efficiently, compared to other network levels. Several proposals to define a “fast restoration” framework have been proposed in [SHAR-00] and [HUAN-00].

Another important aspect to developing a fault management system, is how Backup LSPs could be created and routed. This creation and routing could be done in different ways: in a static manner (pre-establishing LSPs backups until a failure occurs) or dynamically (routing an LSPs backups as an action to recover traffic over the broken working path). Whichever way, alternative routes to recover traffic as fail occurs involve specific routing protocol knowledge. Several proposals to route MPLS LSPs guarantying certain QoS guarantees have been proposed in [KODI-00], [KAR-00] and [SURI-00]. These proposals make use of MPLS capabilities to develop an on-line routing mechanism that allows certain QoS guarantees with the least number of LSP requests rejected.

2.1.1.- MPLS fault control mechanisms.

All protection systems use to follow an application cycle. This cycle starts with a fall detection and ends with the traffic and the working path being recovered. All the phases implicated in this recovery cycle, involve developing components as phases in which that process could be divided. The components firstly needed are: a method for selecting the working and protection paths and a method for bandwidth reservation in the working and protection paths. Once the paths are created a method for signaling the setup of the working and protection paths is required. A fault detection mechanism to detect faults along a path and a fault notification mechanism are necessary to convey information about the occurrence of a fault to a network entity responsible for reacting to the fault and taking appropriate corrective action. Finally, a switchover mechanism to move traffic over from the working path to the protection path is also provided. Optionally, a repair detection mechanism is set up, to detect that a fault along a path has been already repaired. Also a switchback or restoration mechanism, for switching traffic back to the original working path, once it is discovered that the fault has been corrected, is optionally provided.

These are the usual components for a single fault management method. Any protection algorithm involves a definition of each component’s features and behaviors. In this proposal we introduce a new component for selecting and activating each specific component to initiate a specific protection mechanism. This new object triggers the function of every component to activate the fault management mechanism selected.

The development of each MPLS protection component could be constrained by using some features of the MPLS domain. In this section we introduce specific characteristics of MPLS fault management components.

One important aspect is the fault notification method. MPLS lower layers, such as SONET/SDH or the optical layer, have some limitations in covering both notifications (node faults and link faults) [KODI-00]. MPLS allows capabilities which detect link and node faults. The MPLS layer provides the capability for detecting node faults via an appropriately implemented Liveness Message (for example, the “LDP Liveness message”), or via a “Path Continuity Test”. Another capability is that of detecting node misconfigurations. MPLS layers are able to detect node or software misconfigurations by counting errors or corrupted packets, which may be identified by looking at the MPLS label ie: by counting TTL errors or label mismatches.

Independent of the fault indication mechanism signals for indicating a failure (node or link failures), and the signal for the original working path restoration, are: the Failure Indication Signal (FIS) and the Failure Recovery Signal (FRS), which are commonly used by MPLS fault management methods.

These notification methods involve an RNT (Reverse Notification Tree), to indicate the fault to the ingress node or the PSL (Protection Switch Label switch router) [HUAN-00]. PSL are nodes that have the function of switching protected traffic from the working path to the corresponding backup path.

Another aspect is the number of backup LSPs for a protection domain. Setting up a backup LSP for the working LSP is the common way to achieve reliability in MPLS networks. A common solution is to find two disjoint paths. However, this requires, at least, twice the amount of network resources. To overcome this drawback, links on the backup path can be shared between different working paths in a way that single link failure restoration is guaranteed [HASK-00].

One aspect that distinguishes MPLS from other mechanisms is at the level, where protection is applied. In MPLS domains, local repair level or a path repair level is provided. In path level repair, protection is always activated at the edges of the LSP, irrespective of where abouts on the working path the failure occurs. This method should propagate the FIS signal back to the source (Ingress Node), which can be costly, in terms of time. In local repair, protection is activated by an LSR with PSL function along the path to a PML (Path Merge LSR), which merges their traffic into a single outgoing LSP. This method presents the added complication of having to configure multiple backup segments (wherever protection is required), and whenever these resources are reserved “a priori” (and not used) this could result in an inefficient use of resources.

According to the MPLS fault management framework [SHAR-00] a PSL is the transmitter for both the working path traffic and its corresponding backup path traffic. A PSL is the origin of the backup, but does not necessarily have to be an Ingress Node. A PML is the LSR that receives both working path traffic and its corresponding backup path traffic, and merges their traffic into a single outgoing path. This PML may or may not be an Egress Node.

Finally, one aspect, which is not very often discussed, is bandwidth reservation. Algorithms for the problem of setting up bandwidth LSP backups involve information knowledge of network scenario. Depending on the information available we could develop a more or less accurate method. A proposal, which takes up this idea, to develop a bandwidth reservation solution in an MPLS domain with shared backup is introduced in [KODI-00]. In this paper we do not take into account bandwidth reservation considerations.

2.1.2.- MPLS on-line routing algorithms.

A dynamic multilevel protection means that protection scenarios could not be created as a priori. Backup requests arrive one by one, and a priori protection level does not have to be knowledge. Protection environments will be created in a dynamic manner, depending on the needs and features of the protected environment. So, to develop this protection an on-line routing algorithm has to be applied, not only to route new LSP backups, but also LSP working paths.

These kinds of algorithm pre-suppose that new requests arrive one by one and, a priori, there is no knowledge of future demands. Route requests are constrained by the on-line routing algorithm (such a CAC behavior) based on functionality statistics or the type of streams carried by each LSP, to generate more or less protection environments (more or less LSP backups could be routed).

On-line routing algorithms with certain QoS conditions (usually bandwidth guarantees), have been proposed in [KODI-00], [KAR-00] and [SURI-00]. In these proposals, routing algorithms beyond simple WSP (widest-shortest path algorithm) [GUER-97], where MPLS networks features are not taken into account, have been developed. The mechanisms proposed are based on certain MPLS topology parameters: ingress-egress nodes that are known as a priori. With this knowledge that allows where new LSPs requests could be done, and with certain knowledge added (depending on the proposal: residual capacity, link bandwidth or traffic-profiles) a weighted graph is designed where a posteriori "shortest path algorithm" is applied.

2.2.- Thesis Proposal

2.2.1.- A proposal for a dynamic multilevel MPLS fault management

We propose to develop a dynamic multilevel fault management approach. This goal can be achieved gradually. As the backup paths (single backup, segment backups, reverse backups) are being created an available fault management mechanisms table is updated. Based on this table, the decision as to which method has to be activated is taken, according to a pre-defined policy or based on the actual network streams (EXPerimental MPLS header field).
As soon as backups are complete the PSL / PML function, to the nodes that allows the creation of a specific mechanism, could be activated. If more than one method is available, the activation of one of these methods is possible by activating or deactivating the necessary PSLs or PMLs. For example, nodes 1 and 5 (fig. 21) as a PSL and PML, respectively, a centralized recovery method starts. If only nodes 3 and 7 are activated, a local method will be activated.
Finally if nodes 3 and 1 (PSL, PML) are activated, the traffic recovers back to the Ingress Node. Within this backup activation the notification mode should also be activated (see Table 1).

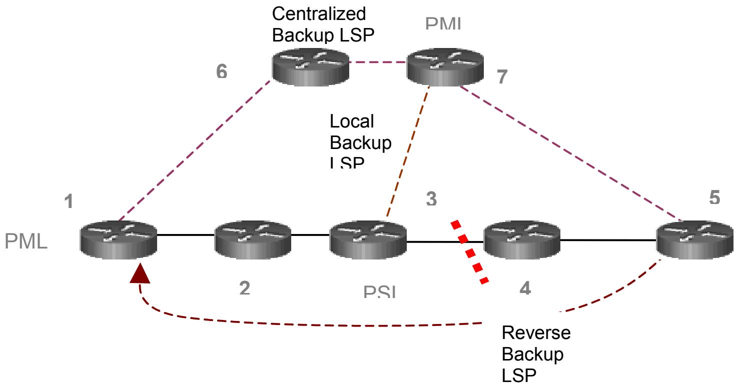


Figure 21 : Complete MPLS protection scenario

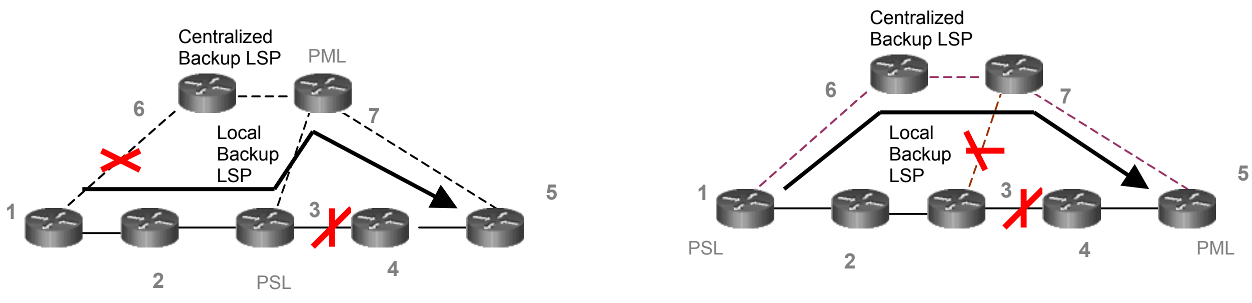
Fault management Method	ACTIVE	PSL	PML	NOTIFICATION METHOD
Centralized	Yes	1	5	RNT
Local	No	1	3	Local
Local	Yes	3	7	Local
Reverse Backup	No	3	1	RNT

Time

Table 3: Table of Fault Management Methods Available

In network scenarios with a high degree of protection requirements, the possibility of a multilevel fault management application could improve performance, compared to the single method application. Nonetheless, complete scenario construction is highly costly (in terms of time and resources), so intermediate scenarios could be built instead. For example our protected domain could start with just a centralized method, and as the protection requirements grow (a node falls repeatedly), a new local backup could be provided, thus making available a new protection mechanisms. These two methods can be activated at the same time. If a fault is located at node 4 or link 3-4, the local method will be applied, transparent to Ingress Node (due to local notification method).

Another advantage of using multilevel protection domains occurs in scenarios with multiple faults. For example, In the figure 22-a if node 4 falls (or LSPs 3-4 or 4-5 faults) and only a centralized backup LSP 1-2-7-5 is used and, afterwards, node 6 or links 1-6, 6-7 fall (during restoration) traffic could be routed to 1-2-3-7-5 avoiding links and node faults. Another example (fig. 22-b) occurs when applying local restoration and link 3-7 falls. In this case, if another backup mechanism (centralized model) is applied the faults are avoided.



Figures 22 (a), (b) : Multilevel protection application.

2.2.2.- Enhancing On-line routing and MPLS protection

Clearly the relationship between on-line routing and MPLS protection mechanisms is how LSPs could be routed. On-line routing mechanism proposes an intelligent way to route new LSPs maintaining certain QoS parameters. In the protection case, a backup LSP need not be routed a priori, even when almost all backups are pre-routed, topological changes or new resource requirements are forced to pick new working LSPs and their corresponding backups. In a multilevel protection proposal the fact is more evident, different protection level involves routing different types of backups, and probably, at different time scales (dynamically).

MIRA and Traffic-profile have included many comments about how on-line routing algorithm affect backup establishment. Only the advantages of using an QoS routing, such as minimizing rejected request or maximizing resource utilization, are taken into consideration in the protection case. One aspect not taken into account is the use of PSL and PMLs in MPLS fault management. MIRA and Traffic-profile propose to use ingress-egress node knowledge to enhance their routing algorithms, but LSPs could be created along a PSL to a PML node, too. A further aspect to expand these proposals to the MPLS protection case is to make use of the Traffic-profile concept to characterize Fall-profiles. The Fall-profiles concept means the probability or the sensibility of a Traffic-profile in the case of fail. Several types of traffic are more sensitive to loses or restoration times than others (this concept is more explained in [OWEN-00]).

In conclusion a review of these algorithms to take into more detailed consideration MPLS fault management, is necessary. Whatever way, on-line routing methods are necessary to develop a fault management system that take in consideration QoS apects.

2.2.3.- Conclusions

In this proposal several on-line routing methods together with MPLS protection methods are introduced. Usual fault management methods pre-establish backup paths to recover the traffic after a failure. MPLS allows the creation of different backup types due to is own characteristics, likewise MPLS is a suitable method to support traffic engineered networks. An introduction of several LSP (Label Switch Paths) backup types and their pros and cons are pointed out. The creation of an LSP involves a routing phase, where QoS aspects, (to achieve a reliable network), must be applied. In a similar way the establishment of LSP backups must be routed under a QoS routing method. When LSP creation request arrives one by one (meaning a dynamic network environment), on-line routing methods are applied. Obviously the relationship between MPLS fault management methods, especially in the creation of LSP backups and QoS on-line routing methods to route new LSP requests, is unavoidable. Both aspects are overviewed in this proposal.

2.3.- Planning and future work

The aim of this project is to demonstrate that applying multilevel protection with on-line based routing allows the achievement of highly reliable networks. To demonstrate this fully more detailed proposal has to be developed. Firstly to prove the argument that a multilevel protection environment allows for much better performance than classical one-level (one backup) proposals, in terms of delay, packet loses and response times. Another important aspect to take into consideration over a multilevel protection is the scalability feature. A second phase is to enhance an on-line based routing proposal taking into account the specific multilevel protection proposal, in a more complete scenario such as the KL-graph (See fig. 20). Finally, a complete MPLS dynamic multilevel protection applying MPLS QoS routing algorithms should be implemented.

References

[ANDE-99] "Requirement Framework for Fast Re-route with MPLS", L. Andersson, B. Cain, B. Jamoussi, (work in progress) Internet Draft draft-andersson-reroute-frmwrk, Oct 1999

[AUKI-00] "RATES : A server for MPLS Traffic Engineering" P. Aukia, M. Kodialam, P. V. N. Koppol, T.V. Lakshman, H. Sarin, and B. Suter. IEEE Network pag. 34 to 41. March/April 2000

[AWDU-00] "A Framework for Internet Traffic Engineering", D.O. Awduche, A. Chiu, A. Elwalid, I. Widjaja, X. Xiao, (work in progress) Internet Draft, draft-ietf-tewg-framework-04.txt, April 2000

[AWDU-99] "Requirements for Traffic Engineering Over MPLS" D. Awduche, J. Malcolm, J. Agogbua, M. O'Dell, J. McManus, Sep 1999 RFC2702

[AWDU-99a] "Extensions to RSVP for LSP Tunnels", D.O.Awduche, L. Berger, D. Gan, T. Li, G. Swallow, V. Srinivasan (work in progress) Internet Draft draft-ietf-mpls-rsvp-lsp-tunnel , Sept 1999.

[AWDU-99b] "Extensions to RSVP for LSP Tunnels", D.O.Awduche, L. Berger, D. Gan, T. Li, G. Swallow, V. Srinivasan, (work in progress) Internet Draft draft-ietf-mpls-rsvp-lsp-tunnel, Sept 1999.

[AWDU-99c] "MPLS and traffic engineering in IP networks" D. O. Awduche, IEEE Communication Magazine, pp. 42--47, December 1999.

[AWDU-99d] "MultiProtocol Lambda Switching: Combining MPLS Traffic Engineering Control With Optical Crossconnects", D. Awduche, Y. Rekhter, J. Drake, R. Coltun, ,(Work in Progress), Internet Draft draft-awduchempls -te-optical-01.txt, Nov 1999.

[CALL-00] "A dynamic multilevel MPLS protection domain" E. Calle, T. Jové, P. Vilà, J.L. Marzo. To appear in DCRN'2001. Third International Workshop on Design of Reliable Communication Networks. 7-10 October 2001, Budapest, Hungary

[CHEN-99] "Reliable Services in MPLS " T. M. Chen and T. H. Oh. December 1999. IEEE Communications. Vol. 37. No 12.

[DAVI-00] "MPLS: Technology and applications", Bruce Davie, Yakov Rekhter, The Morgan Kauffmann, ISBN 1-55860-656-4, May 2000

[DAVI-01] "MPLS using LDP and ATM VC Switching", B. Davie, J. Lawrence, K. McCloghrie, Y. Rekhter, E. Rosen, G. Swallow, P. Doolan, Jan 2001, RFC3035

[GHAN-99] "Traffic Engineering Standards in IP Netrworks Using MPLS", A. Ghanwani, B. Jamoussi, D. Fedyk, P. Ashwood-Smith, N. Feldman, IEEE Commun. Mag., Dec. 1999, vol. 37, no 12., pp. 49-53.

[GUER-97] "QoS Routing Mechanisms and OSPF Extensions", R. Guerin, D. Williams, A. Orda, Proceedings of Globecom 1997.

[HASK-00] "A Method for Setting an Alternative Label Switched Paths to Handle Fast Reroute", D. Haskin, R. Krishnan (work in progress) Internet Draft draft-haskin-mpls-fast-reroute, Nov 2000

[HUAN-00] "A Path Protection/Restoration Mechanism for MPLS Networks", C. Huang, V. Sharma, S.Makam, K. Owens (work in progress) Internet Draft draft-chang-mpls-path-protection, Jul 2000

[JAMI-98] "MPLS VPN Architecture", D. Jamieson, B. Jamoussi, G. Wright, P. Beaubien, (work in progress) Internet Draft draft-jamieson-mpls-vpn-00.txt, Aug 1998

[KAR-00] "Minimum Interference Routing of Bandwidth Guaranteed Tunnels with MPLS TE Aplications (MIRA)" K. Kar, M. Kodialam, T.V. Lakshman. Proceedings of the Conference on Computer Communications (IEEE Infocom), Mar. 2000.

[KINI-00] "Shared backup Label Switched Path restoration", S. Kini, M. Kodialam, T.V. Lakshman, C. Villamizar (work in progress) Internet Draft draft-kini-restoration-shared-backup, Oct 2000

[KODI-00] "Dynamic routing of bandwidth guaranteed tunnels with restoration", M Kodialam and T. V. Lakshman, in Proceedings of the Conference on Computer Communications (IEEE Infocom), Mar. 2000.

[KODI-99] "On-line Routing of Guaranteed Bandwidth Tunnels", M. Kodialam, T.V. Lakshman, Seventh IFIP Workshop on Performance Modelling and Evaluation of ATM/IP Networks, June 1999

[KRIS-99] "Extensions to RSVP to Handle Establishment of Alternate Label-Switched-Paths for Fast Reroute", R. Krishnan, D. Haskin, (work in progress), Internet Draft draft-krishnan-mpls-reroute-resvpext-00.txt, June1999.

[LI-00] "General Considerations for Bandwidth Reservation in Protection", Li Mo. (work in progress) Internet Draft draft-mo-mpls-protection, Jul 2000

[MA-97] "On Path Selection for Traffic with Bandwidth Guarantees", Q. Ma and P. Steenkiste Proceedings of IEEE International Conference of Network Protocols. Oct. 1997

[MAKA-99] "Protection/Restoration of MPLS Networks", S. Makam, V. Sharma, K. Owens, C. Huang (work in progress) Internet Draft draft-makam-mpls-protection, Oct 1999

[MOY-98] "OSPF Version 2",Moy, J, RFC 2328, April 1998.

[MUTH-00] "Core MPLS IP VPN Architecture", K. Muthukrishnan, A. Malis, RFC2917, Sep 2000

[OHBA-01] "MPLS Loop Prevention Mechanism", Y. Ohba, Y. Katsube, E. Rosen, P. Doolan, RFC3063, Feb 2001

[OHBA-99] "Issues on Loop Prevention in MPLS Networks" Y. Ohba IEEE Communications. Dec. 1999 Vol 37, No 12

[OWEN-00] "Network Survivability Considerations for Traffic Engineered IP Networks", K. Owens, V. Sharma (work in progress) Internet Draft draft-owens-te-network-survivability, March 2000.

[REKH-95] "A Border Gateway Protocol 4 (BGP-4)",Y. Rekhter and T. Li RFC 1771,DDN Network Information Center, March 1995

[ROSE-00] "Multiprotocol Label Switching Architecture",E. Rosen, A. Viswanathan, R.Callon, RFC3031,Jan 2001

[ROSE-99] "BGP/MPLS VPNs", E. Rosen, Y. Rekhter, Mar 1999, RFC2547

[SHAR-00] "Framework for MPLS-Based Recovery", V. Sharma, B.M. Crane, S. Makam, K. Owens,C. Huang, F. Hellstrand, J. Weil, L. Andersson, B. Jamoussi, B. Cain, S. Civanlar, A. Chiu (work in progress) Internet Draft draft-ietf-mpls-recovery-frmwrk, Sep 2000

[SHEW-99] "Fast Restoration of MPLS Label Switched Paths", S. Shew,(work in progress) Internet Draft draft-shew-lsp-restoration, Oct 1999

[SURI-00] “Profile-Based Routing: A new Framework for MPLS TE” S.Suri, M. Waldvogel, P.Ramesh Warkhede. ” Washington University Computer Science Technical Report WUCS-00-21, July 2000.

[SWAL-99] "MPLS advantages for Traffic Engineering", G. Swallow, IEEE Commun. Mag., Dec. 1999,vol. 37, no 12., pp. 54-57.

[WANG-96] “Quality of service routing for supporting multimedia applications," Z. Wang and J. Crowcroft, IEEE Journal on Selected Areas in Communications, vol. 14, pp. 1228-1234, Sept. 1996.

[XIAO-00] "Traffic engineering with MPLS in the Internet", X. Xiao, A. Hannan, B. Bailey, and L. Ni, IEEE Network Magazine, March 2000.

Figures

Tables

Appendix : Paper published in DCRN‘2001

The following will appear in the conference “Design of Communications Reliable Networks (DCRN’2001) “, Budapest, Hungary, 7-10 October 2001.

“A dynamic multilevel MPLS protection domain”

Eusebi Calle, Teo Jové, Pere Vilà, Josep Lluís Marzo *
Institut d’Informàtica i aplicacions (IiiA). Universitat de Girona.
Avda. Lluís Santaló s/n, 17071 Girona (SPAIN)
Phone: +34972418475 , Fax: +34972418098
email : {eusebi | teo | perev | marzo} @ eia.udg.es

Abstract

MPLS can be used to support advanced survivability requirements and to enhance the reliability of IP networks. MPLS networks have the capability to establish Label Switched Paths LSPs (similar to the Virtual Circuits concept). This allows MPLS domains to pre-establish protection LSPs, backups for the working LSPs, and achieve better protection switching times than classic IP protection methods.

Several methods for MPLS fault management have been proposed in recent IETF drafts [2], [3], [4], but how to select a method depending on the network scenario has not yet been sufficiently discussed. In this paper we analyze different fault management methods and network scenarios and describe its pros and cons. Our proposal is the progressive creation of a MPLS protection domain. In this domain, different fault management mechanisms are applied, as and they become available. The application of these mechanisms depends on the network status and its protection requirements (protection level).

Keywords: IP and MPLS, Protection and Restoration Algorithms.

Introduction

Protection methods follow a cycle, when the fault is identified until the working LSP is recovered. This cycle involves the development of various components: a method for selecting the working and protection paths and a method for bandwidth reservation in the working and protection paths. Once the paths are created a method for signaling the setup of the working and protection paths is required. A fault detection mechanism to detect faults along a path and a fault notification mechanism are necessary to convey information about the occurrence of a fault to a network entity responsible for reacting to the fault and taking appropriate corrective action. Finally, a switchover mechanism to move traffic over from the working path to the protection path is also provided. Optionally, a repair detection mechanism is set up, to detect that a fault along a path has been already repaired. Also a switchback or restoration mechanism, for switching traffic back to the original working path, once it is discovered that the fault has been corrected, is optionally provided.

These are the usual components for a single fault management method. Any protection algorithm involves a definition of each component’s features and behaviors. In this paper we introduce a new component for selecting and activating each specific component to start a specific protection mechanism. This new object triggers the function of every component to activate the fault management mechanism selected.

In the first section we introduce some features and topics related to fault management components. The next section describes three fault management methods and their pros and cons. Finally, in sections three and four, a completed fault protection scenario is presented. We propose a progressive method for constructing and selecting the optimum mechanism depending on the network status and its protection requirements (protection level).

I.- MPLS Protection environment

The development of each MPLS protection component could be constrained by using some features of the MPLS domain. In this section we introduce specific characteristics of MPLS fault management components.

One important aspect is the fault notification method. MPLS lower layers, such as SONET/SDH or the optical layer, have some limitations in covering both notifications (node faults and link faults) [7]. MPLS allows capabilities which detect link and node faults. The MPLS layer provides the capability for detecting node faults via an appropriately implemented Liveness Message (for example, the “LDP Liveness message”), or via a “Path Continuity Test”. Another capability is that of detecting node misconfigurations. MPLS layers are able to detect node or software misconfigurations by counting errors or corrupted packets, which may be identified by looking at the MPLS label: by counting TTL errors or label mismatches.

Independent to the fault indication mechanism signals for indicating a failure (node or link failures), and the signal for the original working path restoration, are: the Failure Indication Signal (FIS) and the Failure Recovery Signal (FRS), which are commonly used by MPLS fault management methods.

These notification methods involve an RNT (Reverse Notification Tree), to indicate the fault to the ingress node or the PSL (Protection Switch Label switch router) [2]. PSL are nodes that have the function of switching protected traffic from the working path to the corresponding backup path.

Another aspect is the number of backup LSPs for a protection domain. Setting up a backup LSP for the working LSP is the common way to achieve reliability in MPLS networks. A common solution is to find two disjoint paths. However, this requires, at least, twice the amount of network resources. To overcome this drawback, links on the backup path can be shared between different working paths in a way that single link failure restoration is guaranteed [4].

One aspect that distinguished MPLS from other mechanisms is the level, where protection is applied. In MPLS domains, local repair level or a path repair level are provided. In path level repair, protection is always activated at the edges of the LSP, irrespective of where about on the working path the failure occurs. This method should propagate the FIS signal back to the source (Ingress Node), which can be costly, in terms of time. In local repair, protection is activated by an LSR with PSL function along the path to a PML (Path Merge LSR), which merges their traffic into a single outgoing LSP. This method presents the added complication of having to configure multiple backup segments (wherever protection is required), and whenever these resources are reserved “a priori” (and not used) this could result in an inefficient use of resources.

According to the MPLS fault management framework [1] a PSL is the transmitter for both the working path traffic and its corresponding backup path traffic. A PSL is the origin of the backup, but does not necessarily have to be an Ingress Node. A PML is the LSR that receives both working path traffic and its corresponding backup path traffic, and merges their traffic into a single outgoing path. This PML may or may not be an Egress Node.

Finally, one aspect, which is not very often discussed, is bandwidth reservation. Algorithms for the problem of setting up bandwidth LSP backups involve information knowledge of network scenario. Depending on the information available we could develop a more or less accurate method. A proposal, which takes up this idea, to develop a bandwidth reservation solution in an MPLS domain with shared backup is introduced in [7]. In this paper we do not take into account bandwidth reservation considerations.

II.-Main MPLS fault management methods

In this section, three fault management algorithms and their pros and cons are introduced. The following section concludes with a multilevel MPLS protection scenario that covers main features of methods revised in this section.

Centralized model

In this model, an Ingress Node is responsible for resolving the restoration as the FIS arrives. This method needs an alternate disjoint backup path for each active path (working path).

Protection is always activated at the Ingress Node, irrespective of where along the working path a failure occurs. This means that failure information has to be propagated all the way back to the source node before a protection switch is activated. If no reverse LSP is created the fault indication can only be activated as a Path Continuity Test.

This method has the advantage of setting up only one backup path per working path, and is a centralized protection method, which means only one LSR, has to be provided with PSL functions. On the other hand this method has an elevated cost (in terms of time), especially if a Path Continuity Test is used as a fault indication method. If we want to use an RNT as a fault indication method we have to provide a new LSP to reverse back the signal to the Ingress Node.

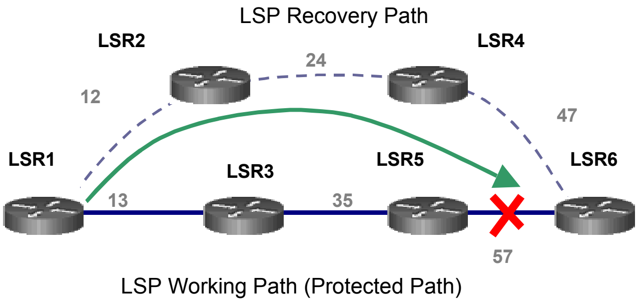


Figure 1 : Centralized model

Figure 1 shows a simple scenario formed by six LSRs where a working path (i.e: LSR1-LSR3-LSR5-LSR6, solid line) and a LSP recovery path (i.e: LSR1-LSR2-LSR4-LSR6, dashed line) are pre-established,. In normal operation, traffic from ingress router LSR1 to egress router LSR6 is carried through LSP working path. When a link fault is detected, (for instance between LSR5 and LRR6), traffic is switched to the LSP Recovery Path, the arrow shows this new path.

LSP segment restoration (local repair)

With this method restoration starts from the point of the failure. It is a local method and is transparent to the Ingress Node. The main advantage is that it offers lower restoration time than the centralized model.

With this method, an added difficulty arises in that every LSR, where protection is required, has to be provided with switchover function (PSL). A PML should be provided too. Another drawback is the maintenance and creation of multiple LSP backups (one per protected domain). This could report low resource utilization and a high development complexity. On the other hand, this method offers transparency to the Ingress Node and faster restoration time than centralized mechanisms.

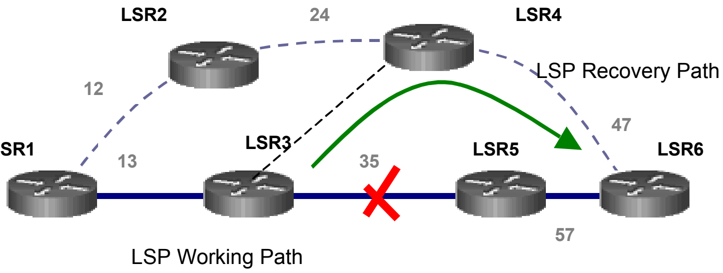


Figure 2 : Local restoration

An intermediate solution could be the establishment of local backup, but only for protection segments where a high degree of reliability is required, supplying only protected path segments.

Figure 2 illustrates this case, the same working path as in centralized model is used (i.e: LSR1-LSR3-LSR5-LSR6, solid line). The LSP recovery path is now formed by LSR3-LSR4-LSR6 that is shorter than the LSP recovery path in the centralized method. When a link failure occurs, traffic is switched from LPD (LSR5-LSR6) which is a segment of the working path to the LSP recovery Path.

Reverse backup

The main idea of this method is to reverse traffic at the point of failure of the protected LSP back to the source switch of the protected path (Ingress Node) via a Reverse Backup LSP.

As soon as a failure along the protected path is detected, the LSR at the ingress of the failed link reroutes incoming traffic by redirecting this traffic into the alternative LSP and traversing the path in the opposite direction to the primary LSP.

This method is especially good in network scenarios where the traffic streams are very sensitive to packet losses. Another advantage is that it simplifies fault indication, since the reverse backup offers, at the same time, a way of transmitting the FIS to the Ingress Node and to the recovery traffic path. One disadvantage could be poor resource utilization. Two backups per protected domain are needed. Another drawback is the time taken to reverse fault indication to the Ingress Node, as with the Centralized model.

Figure 3 shows an example of reverse backup utilization. LSP working and recovery paths are established as in the centralized model, in addition there is a reverse path from LSR5 (LSR5-LSR3-LSR1) which reaches the ingress node. When a link failure is detected in LSP (LSR5-LSR6), the traffic is switched back to LSR1 (ingress node) through the reverse backup LSP, and then carried through the LSP recovery path as in the centralized model.

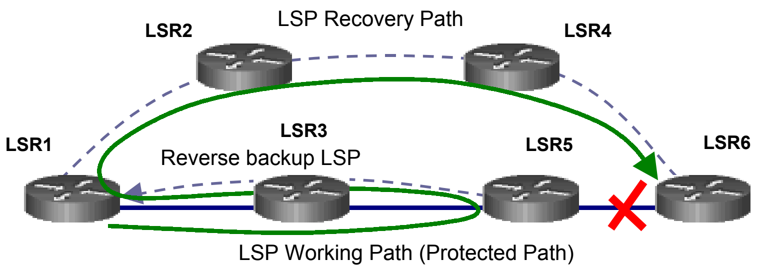


Figure 3 : Reverse backup utilization

III. A proposal for a dynamic multilevel MPLS fault management

We propose to develop a dynamic multilevel fault management approach. This goal can be achieved gradually. As the backup paths (single backup, segment backups, reverse backups) are being created an available fault management mechanisms table is updated. Based on this table, the decision as to which method has to be activated is taken, according to a pre-defined policy or based on the actual network streams (EXPerimental MPLS header field).

As soon as backups are complete the PSL / PML function, to the nodes that allows the creation of a specific mechanism, could be activated. If more than one method is available, the activation of one of these methods is possible by activating or deactivating the necessary PSLs or PMLs. For example, nodes 1 and 5 (fig. 4) as a PSL and PML, respectively, a centralized recovery method starts. If only nodes 3 and 7 are activated, a local method will be activated. Finally if nodes 3 and 1 (PSL, PML) are activated, the traffic recovers back to the Ingress Node. Within this backup activation the notification mode should also be activated (see Table 1).

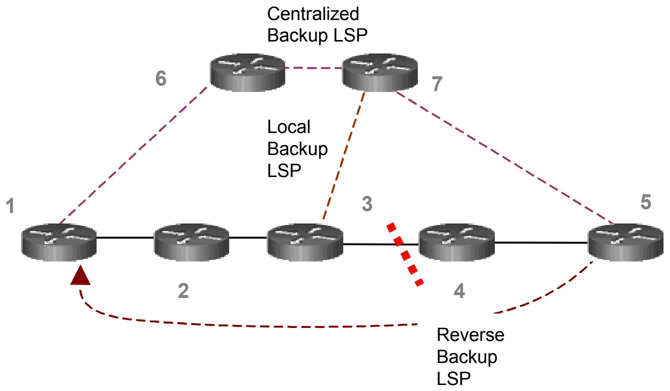


Figure 4 : Complete MPLS protection scenario

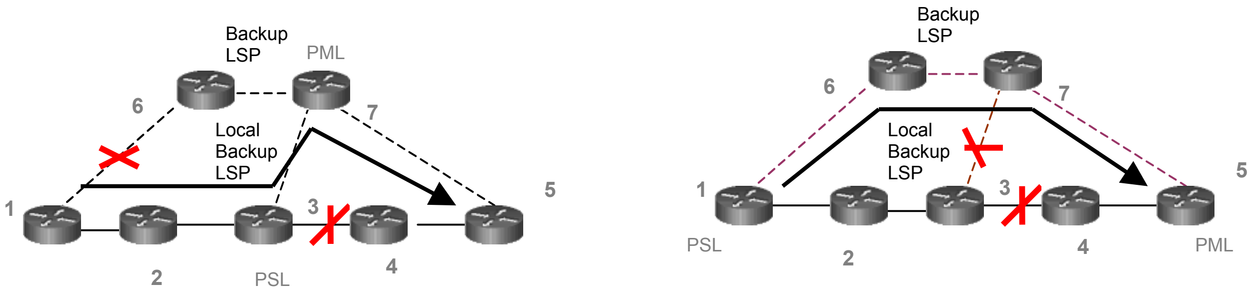
Fault management Method	ACTIVE	PSL	PML	NOTIFICATION METHOD
Centralized	Yes	1	5	RNT
Local	No	1	3	Local
Local	Yes	3	7	Local
Reverse Backup	No	3	1	RNT

Time

Table 1 : Table of Fault Management Methods Available

In network scenarios with a high degree of protection requirements, the possibility of a multilevel fault management application could improve performance, compared to the single method application. Nonetheless, complete scenario construction is highly costly (in terms of time and resources), so intermediate scenarios could be built instead. For example our protected domain could start with just a centralized method, and as the protection requirements grows (a node falls repeatedly), a new local backup could be provided, thus making available a new protection mechanisms. These two methods can be activated at the same time. If a fault is located at node 4 or link 3-4, the local method will be applied, transparent to Ingress Node (due to local notification method).

Another advantage of using multilevel protection domains occurs when in scenarios with multiple faults. For example, (fig 5-a) if node 4 falls (or LSPs 3-4 or 4-5 faults) and only a centralized backup LSP 1-2-7-5 is used and node 6 or links 1-6, 6-7 fall (during restoration) traffic could be route to 1-2-3-7-5 avoiding links and node faults. Another example (fig. 5-b) occurs when applying local restoration and link 3-7 falls. In this case, if another backup mechanism (centralized model) is applied the faults are avoided.

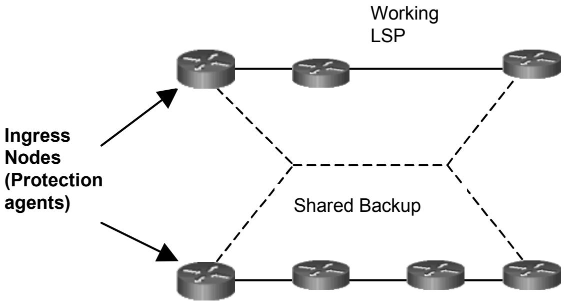


Figures 5 (a), (b) : Multilevel protection application.

IV. Implementation aspects of a dynamic multilevel MPLS protection.

The development of this method could be highly costly (in terms of time and resources). Complete scenario construction could be complex and could report low resource utilization. We propose to analyze network survivability requirements (QoS requirements) and establish different protection levels. Depending on the protection level for a specific MPLS backbone, the development of a more or less complex scenario is constructed.

LSP Backup creation, bandwidth reservation, fault indication, method activation, and PML/PSL functions assignation could be carry out explicitly, via a network administrator, or could be done automatically, via agent application. These agents could be placed on every Ingress Node (see fig. 6) , developing a centralized policy whereby these agents could analyze LSP statistics and network behaviors, and apply defined protection actions.



Ingress Nodes Protection Agent Functions :

•LSP Backup creation.
•Bandwidth reservation
•Fault indication method activation
•Assignation of PML/PSL functions to core LSRs

Figure 6 : Agent application to a dynamic multilevel MPLS protection domain

The specific development the creation and application of agents are beyond the scope of this paper, yet certain proposals, such as [8] could be taken into account when elaborating upon more specific agent development.

Conclusions

In this paper a new component for developing a more specific fault management application is introduced. Progressive construction of a multilevel MPLS protection domain makes available the application of different protection mechanisms. Activation of each method could result in network statistics or in a pre-defined policy.

In network scenarios with a high degree of protection requirements the possibility of a multilevel fault management application could improve performance with respect to single method application. Given that the development of a complete protection domain could be complex and could report low resource utilization, intermediate scenarios can be also built.

Finally, this method could be implemented explicitly, via a network administrator, or automatically, via agent application. More detailed development of this method is a subject for future research.

References

[1] "Framework for MPLS-Based Recovery", V. Sharma, B.M. Crane, S. Makam, K. Owens,C. Huang, F. Hellstrand, J. Weil, L. Andersson, B. Jamoussi, B. Cain, S. Civanlar, A. Chiu, work in progress, Internet Draft draft-ietf-mpls-recovery-frmwrk-00.txt, Sep 2000

[2] "A Path Protection/Restoration Mechanism for MPLS Networks", C. Huang, V. Sharma, S.Makam, K. Owens, , work in progress, Internet Draft draft-chang-mpls-path-protection-02.txt, Jul 2000

[3] "Shared backup Label Switched Path restoration", S. Kini, M. Kodialam, T.V. Lakshman, C. Villamizar, work in progress, Internet Draft draft-kini-restoration-shared-backup-00.txt, Oct 2000

[4]"A Method for Setting an Alternative Label Switched Paths to Handle Fast Reroute", D. Haskin, R. Krishnan, work in progress, Internet Draft draft-haskin-mpls-fast-reroute-05.txt, Nov 2000

[5] “Network Survivability Considerations for Traffic Engineered IP Networks”, K. Owens, V. Sharma.,. work in progress, Internet Draft draft-owens-te-network-survivability-00.txt, March 2000

[6] “General Considerations for Bandwidth Reservation in Protection”, Li Mo. work in progress, Internet Draft draft-mo-mpls-protection-00.txt

[7] “Dynamic Routing of Bandwidth Guaranteed Tunnels with Restoration” M. Kodialam, T.V. Laksman, Inforcom 2000

[8] "A Multi-Agent Approach to Dynamic Virtual Path Management in ATM Networks", P.Vilà, J.L.Marzo, R.Fabregat, D.Harle, Book chapter included in "Agent Technology for Communications Infrastructure", Edited by Alex L.G. Hayzelden and Rachel A. Bourne, © 2001 John Wiley & Sons Ltd, ISBN 0-471-49815-7, pages 167-184