New Approaches to Service Restoration in MPLS-based Networks*

Radim Bartoš, Mythilikanth Raman, and *Arun Gandhi*
Department of Computer Science, University of New Hampshire
Durham, NH 03824, USA. Phone: (603) 862-3792, Fax: (603) 862-3493
E-mail: {rbartos,mraman,agandhi}@cs.unh.edu

Abstract: *MultiProtocol Label Switching (MPLS) has emerged as the technology of choice for future IP networks. As the demands put on the networks increase, MPLS can serve as the basis for providing better reliability, manageability, and overall quality of service. This paper proposes a new approach to service restoration in MPLS networks based on the concept of domain protection, where protection of paths leading from all ingress routers to a common egress router is calculated in one step. The scheme allows a flexible protection path placement independent of the working path placement. Protection paths are calculated using an algorithm with low polynomial complexity. A mechanism for switching over to the protection paths after a failure is detected, that ensures faster switching time, minimal loss and packet reordering, is discussed. A simulation study is used to evaluate the benefits of the proposed scheme and to compare it with two recently proposed schemes for MPLS restoration: RSVP Backup Tunnels and Fast Reroute. The results show that the proposed scheme provides better solution than the Fast Reroute scheme and comparable protection to the RSVP Backup Tunnels while exhibiting lower algorithmic complexity. Unlike the proposed scheme, both Fast Reroute and RSVP Backup Tunnels require simultaneous calculation of working and protection paths in order to guarantee a link disjoint placement.*

1. Introduction

MultiProtocol Label Switching (MPLS) [1] has emerged as one of the technologies that addresses the needs of future IP-based networks. Among the benefits, MPLS promises to improve the routing performance when compared to the standard IP routing, enhances support for traffic engineering, and provides an infrastructure that enables various QoS mechanisms to be implemented in a scalable manner [2]. Finally, MPLS can serve as the layer in which fast, efficient, and network-wide service restoration is provided [3].

Most protocol layers can contribute to the goal of reliable communication. MPLS appears to be the right layer at which high availability of core backbone networks can be ensured. Unlike link and physical layer restoration schemes, such as those in SONET/SDH, an MPLS restoration scheme can take into account the entire network and thus provide more resource efficient solutions. Since MPLS operates on relatively large aggregates of traffic that can be locally switched without a need for network-wide signaling, it is expected that it will provide restoration times significantly shorter than the convergence times of IP routing protocols. Service restoration in MPLS have been extensively studied in the past two years and several schemes have been proposed [4, 5, 6, 7]. This paper describes a new protection scheme for MPLS-based networks and addresses several implementation issues. The long-term goal of this project is to develop technology-neutral architectures, protocols, and procedures for building and operating networks with high availability.

2. Background

Benefits of an MPLS-based protection scheme can be evaluated using many different measures, some of which can be difficult to quantify. There are four main categories of such measures: complexity, resource requirements, restoration times, and the impact of the protection switching on the traffic carried by the network. Complexity of a protection scheme can be measured not only by the algorithmic complexity but also by the overall volume of signaling required to setup protection paths or by the flexibility a network operator has in placing the protection paths. Resource requirements can be measured using the amount of bandwidth allocated to the protection paths, reserved buffer space, or the amount of state information that has to be maintained by the routers. The complete restoration cycle consists of several steps, each of which has its own temporal constraints. The goal of negligible impact on the traffic that is being rerouted is very difficult and costly to achieve. For many applications a certain degree of QoS degradation is acceptable, provided it is balanced by a significantly reduced cost of providing the protection.

There are two schemes for MPLS restoration currently under consideration within IETF, each representing a fundamentally different approach to the problem. The *Fast Reroute* (FR) scheme [8] provides fast restoration times at the expense of comparatively less efficient use of resources. In this scheme a node upon detection of a downstream link failure, reroutes the incoming traffic in the reverse direction of the active path. When the redirected traffic reaches the ingress node, it is switched to a previously established alternate path that is link-disjoint with the active path. Fast protection switching is achieved since the rerouting decision is made using locally available information without a need for signaling. The obvious drawback is the total length of protection paths.

* This research was supported in part by NSF-ARI grant No. 9601602.

The authors of Fast Reroute propose a scheme where an ingress node on detection of traffic in the reverse direction would switch the traffic entering the domain directly to the alternate path. They note that a straightforward implementation of this method would result in temporary packet reordering as packets arriving from the reverse direction are mixed with incoming packets.

RSVP-based Backup Tunnels (BT) [5] take a different approach. A node adjacent to a failed link signals the failure to the upstream nodes. An ingress node upon reception of the failure signal reroutes the traffic over either pre-established or dynamically established path that is link disjoint with the active path. The length of the protection path is less than in the case of the Fast Reroute scheme and the resources are used more efficiently, especially if dynamic protection path establishment is used. However, the traffic, that is in transit between the time the fault is detected and the time when failure signaling reaches the ingress node must be dropped by the node adjacent to the failure. Given the typically high bandwidth-delay product of the current and future backbone networks, potentially large volume of traffic is lost. Furthermore, it ought to be noted that a resource-efficient implementation of this scheme using pre-established protection paths depends on the use of path merging that adds to the algorithmic complexity of the scheme.

3. Proposed Protection Scheme

This paper discusses implementation issues of Two Path protection scheme [6, 7]. The scheme combines the benefits of fast local protection switching with reduced resource requirements while maintaining complexity that is comparable to that of an IP routing protocol such as OSPF. Furthermore, a greater degree of protection path placement flexibility is achieved by utilizing two protection paths placed in a way that no single link failure could cause a simultaneous loss of connectivity between a node and an egress router on both paths. Since protection paths may share links with the working path, the scheme provides greater flexibility than conventional single protection path methods. All protection paths leading towards a common egress router are calculated simultaneously using the following algorithm. A full domain protection is achieved by a concurrent execution of the protection path placement algorithm in each egress router. MPLS traffic engineering methods are then used to establish the protection paths.

3.1 Protection Path Placement

This section presents an algorithm used to place the two sets of protection path as described above. Formal definition of the problem and several alternative approaches can be found in [9, 10]. An earlier version of the algorithm have been presented by the authors in [6, 7].

Input: The MPLS domain with topology represented by graph $G(N, L)$, where N is the set of n nodes, L the set of l links between the nodes, and egress router e. It is assumed that graph $G(N, L)$ has edge connectivity $e_c(G) \geq 2$ thus guaranteeing that the domain can be protected against any single link failure.

Output: Two collections of protection paths connecting ingress routers to egress router e.

Initialization: Find a spanning tree of graph G rooted in the egress router e. Let P be the set of nodes for which the protection paths have been established. Initially it contains the egress router: $P = \{e\}$.

Repeat until all nodes are protected ($P = N$):

1. Select one of the branches of the spanning tree attached to the egress node and mark all its nodes except for the egress node.
2. Scan all marked nodes to find node i that has a link to an unmarked node j.
3. Consider a ring path consisting of the links of the spanning tree leading from e to i, the link between i and j, and the links of the spanning tree between j and e (note that this segment of the ring is empty in the case $j = e$).
4. Place two protection paths along the ring: one in clockwise, the other in counterclockwise direction. The paths originate in the two nodes of the ring that are adjacent to the egress router and follow the ring all the way to the egress node. Merge the created protection paths with the protection paths established in the previous iterations of the algorithm for the protected nodes that are now a part of the egress node. All nodes on the ring are now connected to both protection paths and added to the set of protected nodes P.
5. In the subsequent iteration of the algorithm, consider a new graph constructed by treating all nodes in P as a single node that will act as the egress node and by removing all links that connect two protected nodes.

Options	FR	BT	Proposed
Complexity	Low	High	Low
Path Placement	Restr.	Restr.	Flexible
Restoration Time	Fast	Slow	Fast
Packet Loss	Min.	Yes	Min.
Packet Reordering	Yes	Min.	Min.
Res. Requirements	High	Low	Medium

Table 1: Comparison of MPLS protection schemes.

3.2 Evaluation Criteria

Section 2 outlines some of the general measures that can be used to evaluate an MPLS protection scheme. Table 1 shows a comparison of the three restoration schemes using these criteria.

In this paper we consider two main quantitative measures of quality of a protection scheme: the length of the protection paths and the number of protection paths per link. The length of protection paths was chosen as an indication of the delay the traffic will experience after a link failure. The average of all protection path lengths for the entire domain approximates the overall impact of a fault on the traffic streams. The maximum protection path length gives the worst case scenario, an important measure for real-time traffic provisioning. In addition to the introduced delay, the length of protection paths reflects the amount of resources required to protect the domain.

The number of protection paths that pass through a link is used as an indication of the amount of resources, such as label table sizes and signaling overhead, that are required to setup and maintain the protection. The average and the maximum of the number of protection paths per link are also an indication of how well the protection paths are distributed throughout the network.

3.3 Two Step Protection Mechanism

The signaling of the link failure is done using the recently proposed *Failure Indication Signal* (FIS) [5]. The node connected to a failed link transmits the FIS signal. The FIS propagates hop by hop towards all the ingress routers over a *Reverse Notification Tree* (RNT) which is created along with the working path. Assuming that the working path forms a multipoint-to-point tree with the ingress nodes as the leaves and the egress node as the root, the RNT is the point-to-multipoint tree in the opposite direction.

Switching of traffic over to the protection paths is done in two steps. First, a local decision is made by the nodes, that are adjacent to the failed link, to switch all the working path traffic from the failed link over to the protection paths. Simultaneously, FIS is transmitted along the RNT to inform the affected nodes about the failure. When the nodes on the RNT receive the FIS, they stop transmitting on the working path and the transmission is switched over to the protection paths. This step reduces the length of the protection paths and thus lowers the overall bandwidth utilization. Note that the process of switching may lead to packet reordering if the new path has shorter propagation delay than the old one.

Example in Figure 1(a) shows the utilization of protection paths immediately after the failure of link (A, E). Node E on detection of the fault redirects all traffic destined to egress node A to the protection paths and transmits FIS along the RNT. Upstream nodes upon reception of the failure signal switch over to their protection paths as shown in Figure 1(b).

3.4 Transient Effects of Protection Switching

All the three schemes presented above suffer from transient effects that negatively impact the traffic that is being switched onto the protection paths. Extended periods of *packet loss* could be unacceptable for some streaming applications (e.g., stock market data distribution or multimedia) that typically use unreliable transport protocols. Performance of reliable transport protocols, such as TCP, is also negatively affected by packet loss which often serves as the indicator of the network condition that triggers various congestion control mechanisms leading to reduced application throughput. Similarly, *packet reordering* has an impact on the performance of transport or application protocols. For example, even though most implementations of TCP can handle out of order packet arrivals, some scenarios of out of order delivery may trigger the fast retransmission mechanisms, part of which are steps to avoid congestion that in result reduce the protocol throughput. This work concentrates on reduction of packet loss and reordering during the transitionary periods when traffic is switched to protection paths and proposes a scheme, independent of a particular restoration mechanism, that reduces the negative effects.

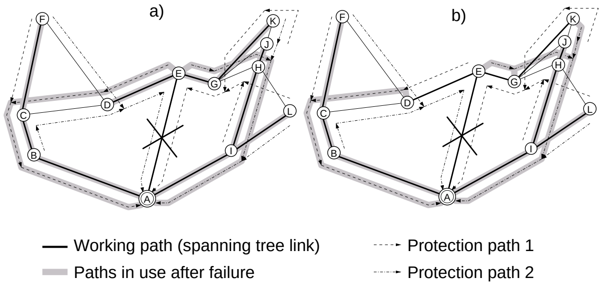

Figure 1: Protection path placement and path utilization (a) immediately after link (A, E) failure and (b) after FISs have propagated.

Packet reordering and loss may occur at a router when the tail of a traffic stream arriving over one path temporarily overlaps with the head of rerouted traffic arriving over protection path. Since the exact propagation delay of a path is typically difficult to predict, such conflicts cannot be resolved during the protection path setup.

There are several straightforward options when dealing with overlapped streams of traffic. Packets from one of the streams can be dropped causing a significant packet loss, the two streams can be merged together causing packet reordering, or packets from the head of the stream can be buffered until all packets from the tail of the other stream are sent on. The last approach does not suffer from reordering, however, the buffers in the router may not be able to store all the packets forcing it to drop the excess ones.

This work addresses the problem by using additional signaling between routers on the protection path. The goal of the signaling is to control the volume of traffic that reaches the merging point by deferring the transmission of packets between routers for as long as possible without overflowing the buffers in the routers. The details of the scheme as well as its performance evaluation cannot be presented here due to the space limitation.

	Backbone 1				Backbone 2			
	TP	ETP	BT	FR	TP	ETP	BT	FR
Maximum number of protection paths	14	13	13	32	13	13	15	35
Average number of protection paths	9.4	9.5	8.7	17.7	9.3	9.3	11.9	17.9
Maximum length of protection paths	5	6	4	7	6	6	5	7
Average length of protection paths	2.2	2.3	2.8	3.5	2.7	2.7	3.1	4.5

Table 2: Simulation results for selected topologies.

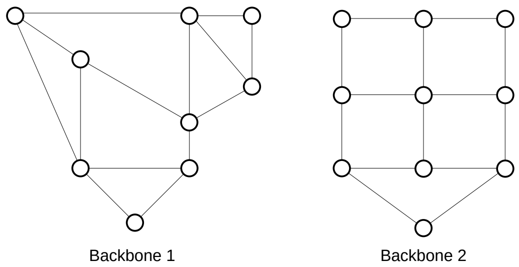

Figure 2: Topologies used in simulations.

4. Performance Evaluation

This section presents comparison of the proposed scheme with the two existing schemes for MPLS protection outlined in Section 2. Two variants of the proposed scheme are considered in the experiments: one which makes use of arbitrary choices at all points of the algorithm except for the spanning tree selection where Dijkstra's algorithm is employed (Two Path - TP). The second variant (Extended Two Path - ETP) utilizes the shortest ring selection heuristic [7]. Performance of the protection schemes was evaluated using the criteria described in the previous section: lengths of protection paths and the number of protection paths per node. For both criteria average and maximum values are considered.

Table 2 summarizes the results of simulation of two backbone network topologies shown in Figure 2. The table show that in almost all case the proposed scheme performs better than FR. When compared to BT scheme, the proposed method provides comparable results while utilizing less complex path placement algorithm. Extended Two Path version of the proposed scheme provides slightly better results than the standard Two Path version.

5. Conclusions

This paper describes a novel approach to service restoration in MPLS-based networks. An algorithm used to place the protection paths is introduced together with a methodology to evaluate protection schemes. Two step protection mechanism is proposed to facilitate fast local protection switching as well as resource-efficient longer-term service restoration. The proposed scheme have been extensively evaluated using simulators. Given the space restriction of the paper, only a small fraction of the results can be reported here.

References

[1] E. C. Rosen, A. Viswanathan, and R. Callon, "Multiprotocol label switching architecture." Work in progress *[draft-ietf-mpls-arch-07.txt]*, July 2000.

[2] G. Swallow, "MPLS advantages for traffic engineering," *IEEE Communications Magazine*, vol. 37, pp. 54–57, December 1999.

[3] V. Sharma et al., "Framework for MPLS-based recovery." Work in progress *[draft-ietf-mpls-recovery-frmwrk-01.txt]*, November 2000.

[4] D. O. Awduche, L. Berger, D. Gan, T. Li, G. Swallow, and V. Srinivasan, "RSVP-TE: Extensions to RSVP for LSP tunnels." Work in progress *[draft-ietf-mpls-rsvp-lsp-tunnel-07.txt]*, August 2000.

[5] S. Makam, V. Sharma, K. Owens, and C. Huang, "A path protection/restoration mechanism for MPLS networks." Work in progress *[draft-chang-mpls-path-protection-02.txt]*, November 2000.

[6] R. Bartoš and M. Raman, "A scheme for fast restoration in MPLS networks," in *Proc. of the Twelfth IASTED International Conference on Parallel and Distributed Computing and Systems (PDCS), Las Vegas, NE* (M. Guizani and X. Shen, eds.), pp. 488–493, November 2000.

[7] R. Bartoš and M. Raman, "A heuristic approach to service restoration in MPLS networks." To appear in Proc. of the 2001 IEEE International Conference on Communications (ICC), Helsinki, Finland, June 2001.

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

[9] A. Itai and M. Rodeh, "The multi-tree approach to reliability in distributed networks," *Information and Computation*, vol. 79, pp. 43–59, October 1988.

[10] M. Médard, S. G. Finn, R. A. Barry, and R. G. Gallager, "Redundant trees for preplanned recovery in arbitrary vertex-redundant or edge-redundant graphs," *IEEE/ACM Trans. on Networking*, vol. 7, pp. 641–652, October 1999.