Adaptive Segment Path Restoration (ASPR) in MPLS Networks

Song Dong and Chris Phillips
Department of Electronic Engineering, Queen Mary, University of London
London, E1 4NS, UK Phone: (20) 7882-5356, Fax: (20) 7882-7997
Email: {song.dong, chris.Phillips}@elec.qmul.ac.uk

Abstract: Two important factors must be considered when selecting a restoration scheme in MPLS networks. Firstly, the restoration time, and consequentially the packet latency of the restored traffic, has to satisfy the requirements of real-time services. Secondly, the spare capacity requirement should be cost-effective. This paper proposes a novel proactive restoration scheme that is both fast-acting and resource efficient, called Adaptive Segment Path Restoration (ASPR). The basic idea is to divide a LSP into several segments according to the network topology. For each segment of the primary path, a backup path is provided. A comparative study of ASPR is provided which shows that of the schemes considered, ASPR has the shortest overall restoration length and smallest backup LSP hop count, whilst remaining better than most other restoration schemes in terms of its resource requirements.

Key words: MPLS, Network Restoration

1. INTRODUCTION

With the migration of real-time and high-priority traffic to IP networks, and with the need for IP networks to increasingly carry mission-critical business data, network resilience has become critical for future Internet networks. Multi-Protocol Label Switching (MPLS) [1], which provides a new technical foundation for the next generation Internet networks introduces a fast signalling scheme for Internet traffic engineering, and is well suited to support various resiliency schemes[2][3].

The original version of this chapter was revised: The copyright line was incorrect. This has been corrected. The Erratum to this chapter is available at DOI: 10.1007/978-0-387-35620-4_43

D. Gaïti et al. (eds.), *Network Control and Engineering for QoS, Security and Mobility*

1.1 Network Resilience Schemes

Network resiliency schemes can be roughly classified in two ways as reactive restoration and preplanned restoration/ protection [4][5]. Reactive restoration dynamically allocates spare resources for the alternate route. It has the advantage of being cost efficient since none resource is allocated before the failure. The drawbacks of this approach are, firstly, that the amount of unreserved resources may not be adequate and some flows may have to be rejected and, secondly, that the recovery latency can be several seconds or even longer, especially in heavily loaded networks [6]. This makes these schemes only suitable for best effort services. Preplanned restoration reserves some resources, identifying backup paths at the time of establishing the primary paths to protect traffic against possible faults. Since these schemes do not need the time-consuming connection reestablishment process, preplanned restoration is capable of restoring traffic within a very short time. A successful application is within SONET/SDH protection rings. Here a restoration time of less than 50 milliseconds has become a benchmark within the industry. However, the drawback of preplanned restoration is the high cost. There will generally be an investment of at least 100% in transmission capacity redundancy [7]. For better resource utilization, resource sharing can be employed [4]. If two primary paths do not fail at the same time, their backup paths can be shared with each other, and thus the costs may be reduced.

Network resilient schemes can also be categorized as link or path restoration[4][5]. Link restoration employs local rerouting, while path restoration uses end-to-end rerouting. Link restoration reroute traffic around the failed component. When a link fails, a new path is selected between the end nodes of the failed link. Link restoration has an advantage of being able to restore traffic in a very short time but it requires setting aside significant spare resources for the backup path. In path restoration, a backup path is established between the end nodes of the failed primary path. This method has better resource utilization than link restoration since it has a better scope for resource sharing. However, the notification of the fault to the ingress node may take a long time, thus making this approach unattractive for the real-time services. Comparative study of these two restoration strategies on spare capacity requirement can be found in [8].

1.2 MPLS Restoration

Two mechanisms have recently been proposed for the restoration of Label Switched Paths (LSP) set up in the MPLS networks, Gan's Scheme [9] and the Haskin's Scheme [3][10].

In Gan's Scheme, extensions to RSVP have been made to incorporate the concept of LSP tunnels into the RSVP flows. Together, these make it possible for routers using RSVP to create detours that can route around downstream links and nodes. As a result, a LSP can quickly and automatically use an alternative by redirecting the user traffic to the pre-computed and pre-established detour routes in event of network link and node failures. The drawback of this method is that it necessitates significant resources to be set aside, as we will show in a later section.

Haskin's scheme is to reverse traffic at the point of the failure back to the ingress node of the protected LSP and redirect it via an alternative pre-configured LSP. This mechanism involves the setting up of two backup paths (separate from the working path). One of these backup paths, called the reverse path, runs in the opposite direction to the working path, from the penultimate node to the ingress node, via the same nodes that are along the working path. The second backup path is established from the ingress node to the egress node via nodes that are path and link disjoint with the working path. When a failure arises, traffic is first redirected along the reverse path to the ingress node and from there it is forwarded along the alterative backup LSP.

2. ADAPTIVE SEGMENT PATH RESTORATION SCHEME

The proposed scheme, entitled Adaptive Segment Path Restoration (ASPR), establishes backup LSPs for a given primary LSP using the standard MPLS signaling protocols. The basic idea is to divide a LSP into several segments. For each segment of the primary path, we provide it with a backup path. The segmentation of the primary path is adaptive to the topology of the network, allowing for more efficient resource usage whilst yielding restoration times comparable to link restoration.

ASPR is performed together with the deployment of working path LSPs and consists of two phases. First, during the propagation of the forward signalling message, the primary path is divided into several segments according to the topology of the network. For each segment a backup path is then calculated. Then, along with the backward signalling message, the segmentation of the primary LSP and backup LSPs are further amended adaptively to the topology of the network. The backup LSPs are deployed only after the primary LSP is established.

Although this scheme does not have a particular requirement on the signaling protocol, and is capable of working with both CR-LDP and RSVP-TE, here we only illustrate it with application using CR-LDP.

2.1 MPLS Traffic Restoration Cycle

MPLS traffic restoration occurs when there is a failure in one or more network components. A whole MPLS traffic restoration cycle includes *Failure Detection*, *Fail Notification* and *Traffic Restoration*. A more detailed discussion about the MPLS restoration cycle can be found in [2]. The restoration time is defined as the interrupted period before the traffic is completely restored. It can be calculated as following:

$$T = t_d + t_n + t_s + t_r \qquad (1)$$

Here t_d is the time between the network link failure and the failure is discovered by the MPLS restoration mechanism. This time may highly depend on lower layer protocols and usually is a constant for a given network.

t_n is the time taken by the notification message to travel from the LSR which detects the failure to the Path Switching LSR (PSL) which takes charge of the traffic switching to the backup LSP. In order to reduce the restoration time during its propagation, the notification message is usually assigned the premier priority during its transmission. After receiving a notification message, an intermediate LSR will forward it immediately without putting it in queue buffers. Thus t_n contains mainly the propagation delay and is proportional to the length of the primary LSP it protects.

$$t_n \propto L_{prim} \qquad (2)$$

t_s is the time consumed by the PSL which takes charge of the traffic switching from the primary LSP to the backup LSP. The operation of switching traffic to the backup LSP and the time it takes vary a lot and depend on the communication layer the MPLS signalling is associated with. However it varies, for different restoration schemes in the same network, we can treat t_s as a constant.

t_r represents the time taken by the diverted traffic travelling along the backup LSP till merging back into the Path Merge LSR (PML) after the failure point. This period is determined by the propagation delay, queuing delay, and processing delay if MPLS restoration is performed in higher layers, while in lower layer the queuing delay is replaced by the time taken for cross connect action of the physical switching unit, such as an OXC in optical layer operation. Compared to propagation delay and queuing delay, the processing delay is very small and can be omitted. Propagation delay is proportional to the length of the backup LSP. Thus,

$$t_r \propto L_{backup} \qquad (3)$$

In advance of the commencement of traffic flows, Service Level Agreements (SLAs) are decided to satisfy the QoS requirement of different services. The traffic flows are treated with different priorities in buffers, typically resulting in experiencing different delays. However, for the same service, the queuing delay can be treated approximately proportional to the number of hops of the backup LSP. Thus,

$$t_r \propto H_{backup} \qquad (4)$$

From (1), (2), (3) and (4), we obtain

$$T \propto L_C, H_{backup} \quad where \quad L_C = L_{prim} + L_{bakcup} \qquad (5)$$

From (5), we can conclude that the total restoration time is proportional to the total length of primary and backup LSP and the hops of backup path.

2.2 Adaptive Segment Path Restoration Algorithm

In order to restore the traffic within the time required by the service whilst making more efficient use of the network's resources at the same time, ASPR divides the primary LSP into several segments. We call the end nodes of each segment the *Segmentation Points*. The segmenting principle is that all the adjacent LSRs that have the same *Restoration Length* are put in the same segment. Then in each segment, a backup path is found to cover possible link failures within this segment. The purpose is to make the *Restoration Length* and backup hops satisfy the QoS requirement of the different services being transported. The segmenting of the LSP is adaptive to the topology of the network and further to QoS requirement of each service.

Assume a MPLS network with a topology represented by the graph $G(V, L)$, where V is the set of v nodes and L is the set of l links between the nodes. Furthermore, assume that graph G is two-connected redundant and therefore can be protected against any single link failure. A LSP P is a sequence $<v_1, l_{12}, v_2, l_{23}, \ldots, v_i, l_{i,i+1}, v_{i+1}, \ldots, v_{k+1}>$, where $l_{i,i+1}$ is a link with endpoints v_i and v_{i+1} (for $i = 1, \ldots, k$), v_1 is the ingress node and v_{k+1} is the egress node. The algorithm attempts to divide the primary LSP into several segments and deploys a backup LSP for each segment, respectively. Given a primary LSP P, *Figure 1* shows how the segmentation point is located.

In the algorithm, *S_P* represents the *Segmentation Point* and is initialised as next hop to the current node on primary LSP. L_C is initialised as the *Restoration Length* of next hop. The algorithm is to find the farthest node to the current node on the primary LSP that has the same *Restoration Length* as that of the anterior nodes.

```
begin SP(G, P)
    $S\_P = v_2$
    $L_{prim} = l_{12}$
    Find backup LSP to $v_2$: $P_b$
    $L_{backup} = Length(P_b)$
    $L_C = L_{prim} + L_{backup}$
    for $i = 2$ to $|P|-1$ do
        $L_{prim} += l_{i,i+1}$
        Find backup LSP to $v_{i+1}$ $P_b$
        $L_{backup} = Length(P_b)$
        if $L_C < L_{prim} + L_{backup}$
            break
        else
            $S\_P = v_{i+1}$
            $L_C = L_{prim} + L_{backup}$
end
```

Figure 1. Segmentation Point Location Procedure

Figure 2(a) shows an example of how a *Segmentation Point* is found and the primary and backup LSPs are calculated.

2.3 Setup the Primary Path

At the ingress LSR, the Segmentation Point Location Procedure is performed to find the next hop *Segmentation Point*. At the same time, the backup LSP to the Segmentation Point is also calculated.

In ASPR, all backup LSPs are deployed only after the primary LSP has been set up. This makes the setting of Label Mapping Table in PSL and PML much easier. When the ingress LSR receives a request to set up primary LSP, the Segmentation Point Location Procedure is performed to find the next hop *Segmentation Point* along with the backup LSP to it. Since it should wait until the deployment of primary LSP is finished, the backup LSP is recorded with the pending Label Request Message of the primary LSP. In the Label Request Message, a new field named *Segmentation Point TLV* is inserted. It denotes the LSR ID of next hop *Segmentation Point.* When the downstream LSRs, except the egress LSR, receive the Label Request Message, the following procedure is performed:

1. The *Segmentation Point TLV* is checked to find if current LSR is the *Segmentation Point.* If not, go to 6.
2. The Segmentation Point Location Procedure is performed to find the next hop *Segmentation Point.*

3. The backup LSP from current LSR to the *Segmentation Point* is calculated.
4. A Label Request Message is produced with the new *Segmentation Point* value and sent to the downstream LSR.
5. The backup LSP is saved with the pending Label Request Message. Exit.
6. A Label Request Message is produced with the *Segmentation Point* and sent to the downstream LSR. Exit.

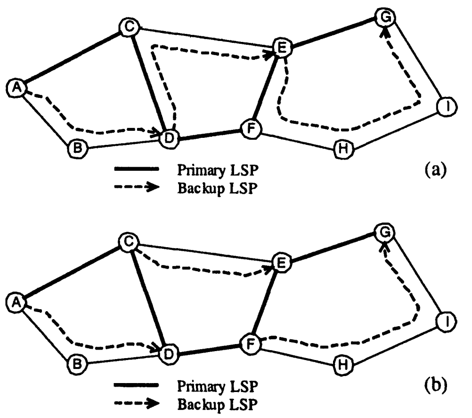


Figure 2. Segment Path Refinement

When the Label Request Message reaches the egress node, a Label Mapping Message is produced to send to the upstream LSR. In order to share the spare resources between different backup LSPs, the links intend to be protected are recorded along the path of Label Mapping Message. In ASPR, a new field named *Primary Segment Path Vector TLV* is created and inserted in the Label Mapping Message during the establishment of the primary LSP. It denotes the LSR ID list of the primary LSP segment. It is further inserted in the Label Request Message during the establishment of the backup LSP in order to realize resource sharing.

2.4 Backup Path Refinement

In most circumstances, *route rewind* will take place for the backup LSPs, which is shown in *Figure* 2(a). Here, LSR *D* and *E* are set as the *Segmentation Points* and backup LSP origins for each segment path are calculated. It is better to set LSR C and F instead of LSR D and E as the PSL for the second and third segments. Thus certain amendment is needed for the backup LSP to avoid *route rewind.* To realize this function, a new optional parameter called *Backup Explicit Route TLV* is inserted in the Label Mapping Message.

When the egress LSR receives the Label Request Message, a Label Mapping Message is produced with the *Primary Segment Path Vector TLV* set as egress LSR ID.

When a LSR receives a Label Mapping Message, the following procedure is performed:

1. The current LSR ID is inserted into the *Primary Segment Path Vector TLV* list.
2. If there is a *Backup Explicit Route TLV* in the received message, go to 11.
3. If there is a backup LSP with the pending Label Request Message, go to 5.
4. A Label Mapping Message is created with the *Backup Explicit Route TLV* and sent to upstream node. Exit.
5. If current LSR is the ingress node, go to 7.
6. If the next hop of the Label Mapping Message and that of the backup LSP have the same LSR, go to 10.
7. The current LSR is the PSL for this segment. It starts to deploy the backup LSP for this segment. The *Primary Segment Path Vector* list is inserted in the Label Request Message to indicate the links it intends to protect.
8. If the current LSR is the ingress node, the deployment of primary LSP and backup LSPs are finished. Exit.
9. A new *Primary Segment Path Vector* list is created with only current LSR ID and is put in the Label Mapping Message, which is sent to upstream node. Exit.
10. A Label Mapping message is created with the *Primary Segment Path Vector* list. The first hop is deleted from the backup LSP, which then is put in the *Backup Explicit Route TLV* in the Label Mapping message. The Label Mapping Message is sent to upstream. Exit.
11. If the next hop of the Label Mapping Message and the backup LSP, denoted by *Backup Explicit Route TLV*, have the same LSR ID then go to 10; otherwise, go to 7.

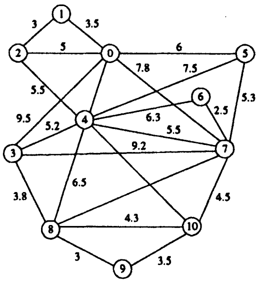
Figure 2(b) shows the result of the amendment of the backup LSPs. Here, link *AC* is protected by backup LSP *ABD*. Link *CD* and *DF* are protected by backup LSP *CE*. Link *FE* and *EG* are protected by backup LSP *FHIG*.

3. PERFORMANCE EVALUATION

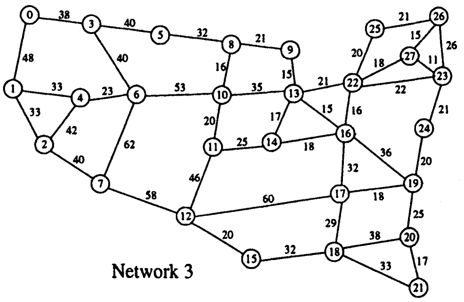
Extensive simulations have been carried out using OPNET™ to analyse the effectiveness of our algorithm in a wide variety of network environments. Four networks shown in *Figure 3* are used in the comparison of different restoration schemes.

The experiment implements the ASPR as well as other MPLS restoration schemes including: Link Restoration, Path Restoration, Gan's Scheme and Haskin's Scheme.

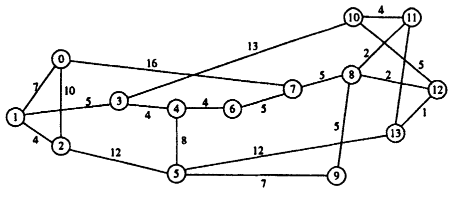
In all test networks, each node sets up 100 primary LSPs, of which the egress nodes are uniformly distributed over the set of network nodes. The bandwidth requirement of each flow is one unit. The shortest path algorithm is used to calculate the primary and backup LSPs.



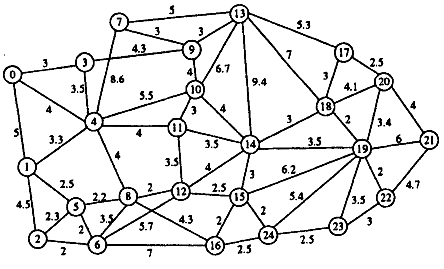
Network 1



Network 3



Network 2



Network 4

Figure 3. Network Topologies

3.1 Spare Capacity Requirement

The Spare Capacity Requirement is used to evaluate the cost efficiency of the restoration schemes. It is defined as the ratio of the total backup resource cost to that of the primary flows. We assume that the link cost is proportional to its length. *Figure 4* shows the results. We observe that Path Restoration has the best performance on Spare Capacity Requirement. Gan's

Scheme has the worst performance and is even worse than that of Link Restoration. ASPR has a better performance than Link Restoration. This is also expected since ASPR uses one backup LSP for all the possible links in the same segment, which results in better resource sharing. Haskin's Scheme has the second best Spare Capacity Requirement performance next to Path Restoration. Although it requires an additional reverse segment, it still provides good resource usage.

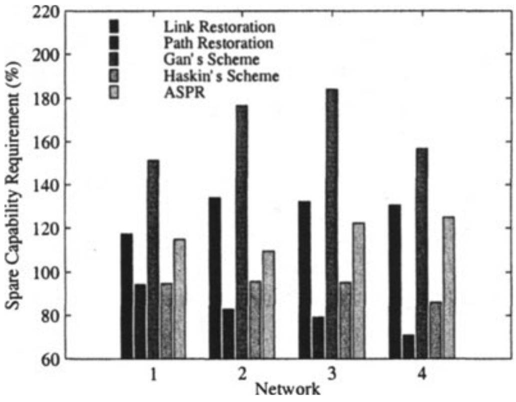


Figure 4. Spare Capacity Requirement

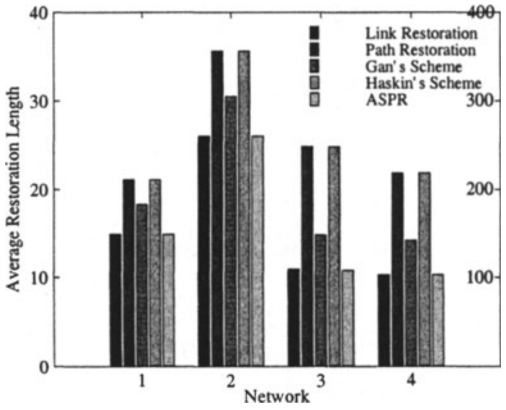


Figure 5. Average Restoration Length

3.2 Restoration Length

Figure 5 shows the average *Restoration Length.* In the diagram, Y label for the Network 1, 2 and 4 are put at the left side whilst those for network 3 are at the right side. We observe that ASPR and Link Restoration have the shortest average *Restoration Length,* nearly half that of Haskin's Scheme and Path Restoration. This means the restoration time of ASPR is the shortest.

3.3 Backup LSP Hops

The restoration time is not only related to the *Restoration Length,* but also to the hop count of the backup LSP. In higher layer restoration, the increased hop count of the backup LSPs, also increases the probability of greater queuing delays, and thus the overall latency of the restoration path. In optical layer restoration, more hops of backup LSP mean more OXCs need to perform a cross connect action to divert the traffic, giving a bigger restoration latency and expense. *Figure 6* shows the average backup LSP hops. We observe that ASPR has the best average backup LSP hops performance. Haskin's Scheme has the poorest performance. Because it is based on the end-to-end restoration, the hops of backup path also depend on the size of the network. Gan's Scheme has a better performance than Haskin's Scheme and Path Restoration but worse performance than ASPR.

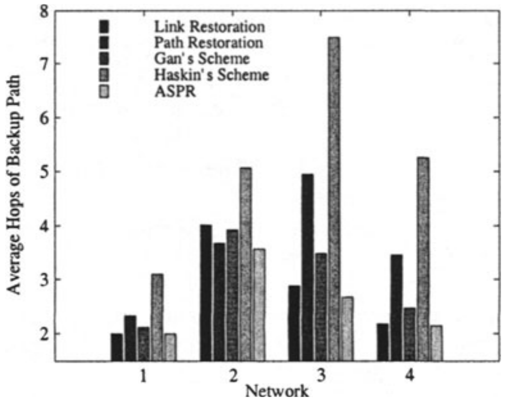


Figure 6. Average Hops of Backup Path

4. CONCLUSION

This paper has proposed a novel MPLS restoration scheme called Adaptive Segment Path Restoration. A comparative study of ASPR with other schemes shows that it compares favourably with the two MPLS restoration schemes currently being considered IETF. Although, Haskin's Scheme has a lower spare capacity requirement, its restoration time performance is the poorest. In addition, the backup path length and hops are typically much larger, yielding latencies that could be unacceptable for the real-time services. Gan's Scheme has a shorter restoration length and smaller backup path hop count than Haskin's Scheme, so providing faster restoration and lower packet latency. However, its spare capacity requirement is the greatest, needing an average 170% redundancy. This would be unacceptable to most service providers. ASPR is shown to have the best restoration performance with the least backup path hop count, giving rise to the lowest packet delay for the restored traffic. These factors are particularly relevant to the transport of real-time services. In addition, it is better than most of the other restoration schemes with regard to the spare capacity requirement.

5. REFERENCES

[1] E.C. Rosen, A. Viswanathan, and R. Callon, "Multiprotocol Label Switching Architecture", IETF RFC 3031, January 2001.

[2] V. Sharma, et al., "Framework for MPLS-based Recovery", Internet Draft draft-ietf-mpls-recovery-frmwrk-03.txt, work in progress, July 2001.

[3] A. Farrel and B. Miller, "Surviving Failures in MPLS Networks", White Paper, Data Connection Limited, http://www.dataconnection.com, Feburary 2001.

[4] C. Dovrolis and P. Ramanathan, "Resource Aggregation for Fault Tolerance in Integrated Service Networks",

[5] G. Mohan and C.S.R. Murthy, "Lightpath Restoration in WDM Optical Networks", IEEE Network, pp. 24-32, November/December 2000.

[6] Y. Ye, S. Dixit, and M. Ali, "On joint protection/restoration in IP-centric DWDM based optical transport networks", IEEE Communications Magazine, vol. 38, Issue: 6, pp. 174-183, June 2000.

[7] D. Stamatelakis and W.D. Grover, "IP Layer Restoration and Network Planning Based on Virtual Protection Cycles", IEEE Journal on Selected Areas in Communications, Vol. 18, No. 10, pp. 1938-1949, October 2000.

[8] Y. Xiong and L.G. Mason, "Restoration Strategies and Spare Capacity Requirements in Self-Healing ATM Networks", IEEE/ACM Transactions on Networking, Vol. 7, No. 1, pp. 98-110, February 1999.

[9] D.-H. Gan, et al., "A Method for MPLS LSP Fast-Reroute Using RSVP Detours", Internet Draft draft-gan-fast-reroute-00.txt, work in progress, April 2001.

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