Sriganesh Kini
- Mahi Networks

Murali Kodialam
T.V.Lakshman
- Bell Labs

Sudipta Sengupta
- Tellium, Inc.

Curtis Villamizar
- Avici Systems

May 2001

Shared Backup Label Switched Path Restoration

draft-kini-restoration-shared-backup-01.txt

Status of this Memo

Abstract

Traffic engineering using MPLS involves the setting up of label switched paths (LSP) possibly with explicit routing and with bandwidth guarantees (for label switched paths). The reliability of these LSPs can be increased by providing a backup LSP onto which traffic can be switched upon failure of an element in the path of the active LSP. Backup LSPs can be routed in a way that bandwidth can be shared between backup links of more than one active path while still

guaranteeing recoverability for a set of failures. This sharing greatly increases the network efficiency, thereby increasing the number of LSPs that can be carried while maintaining guarantees. Algorithms which can route such recoverable LSPs while using only aggregate network usage information are being developed.

This informational draft illustrates the concept of sharing links along backup paths and examines the requirements from link state information and the signaling functions.

1. Introduction

The Multi Protocol Label Switching (MPLS) working group has developed a framework and standards which enable traffic engineering of networks. The framework is described in [8] and the architecture is described in [7]. The MPLS framework is becoming increasing popular to traffic engineer IP networks.

MPLS uses the label swapping paradigm to switch data over an LSP. The functional capabilities required for operations in an MPLS domain are described in [9]. The network layer routing determines the route of an LSP from the topology of the network and the current demands of the applications utilizing the network. Link state routing protocols like OSPF (as described in [1]) and IS-IS (as described in [2]) can provide the topology information to network layer routing that engineers traffic. Signaling protocols like RSVP (as described in [9] and [10]) and LDP (as described in [11] and [12]) are then used to setup the LSP.

Since a LSP traverses a fixed path in the network, its reliability depends on the links and nodes along the path. Traditionally IP networks have carried only best-effort traffic. However new applications are using the IP network infrastructure in ways that make it highly desirable to incorporate faster repair.

MPLS based recovery provides a faster restoration mechanism than layer 3 routing. Several methods have been proposed for MPLS based recovery. A framework and terminology for MPLS based recovery is described in [3].

Setting up a backup LSP for an active LSP (e.g. [6]) is one way to achieve reliability. A straightforward solution to this problem is to find two disjoint paths. However this requires at least twice the amount of network resources. For a restoration objective like single link failure recovery, links on the backup path can be shared between

different active paths in a way that single link failure restoration is guaranteed. This must be done without requiring per-LSP routing information for all LSPs currently c:arried by the network, since keeping track of per-LSP routing information for the whole network is not feasible. It is more desirable to efficiently route recoverable LSPs with shared backups using only aggregate network usage information. Aggregate information useful for setting up shared backup paths is obtainable by adding a few new information elements to the link state advertisement of a link state routing protocol like OSPF or ISIS. Algorithms which can operate using aggregate information are being developed. An example of one such algorithm is described later. Other examples of such algorithms are

in [4],[5]. These algorithms achieve a very high degree of sharing by using aggregate information that can be conveyed by a link state routing protocol.

The key concept of sharing backup paths is described in Section 2. The requirements from link state protocols and signaling protocols is briefly examined in Section 3.

2. Concept of Sharing Backup Paths

A brief description of the concept of sharing is described in this section.

The model under which this principle of sharing is expected to be deployed is very general. The failure entity for which protection is provided could be a link, node, or Shared Risk Link Group (SRLG). Requests for LSPs should be routed as they arrive (online). A new type of LSP can be computed where given a source and destination, an active path and a backup path (possibly shared) is calculated. Bandwidth on links on the backup path are possibly shared between backup paths of other active paths in a way that single link/node/SRLG failure restoration is guaranteed. As requests for setup and teardown of such LSPs arrive and link failures occur, the total link bandwidth allocated for active paths and backup paths vary accordingly.

Sections 2.1 through 2.3 illustrate the sharing concept through some examples and Section 2.4 outlines one simple algorithm that achieves sharing and guarantees single link failure recovery. This algorithm needs only aggregate information. [4] and [5] are other instances of similar algorithms. They achieve a very high degree of sharing using

only aggregate information. Section 2.5 outlines a framework for distributed determination of the shared bandwidth on a link for backup paths during path signaling.

2.1 Sharing Backup: Single Link Failure Recovery

```
                        L1
                   ____________________
                  /                    \
                 |                      |
               ----     L1b   L2b      ----
              | A  |----------------| B  |
               ----                    ----
                 |                      |
                  \____________________/
                        L2
```

Figure 1 : Sharing backup links with link failure recovery

Figure 1 illustrates a simple case of sharing of backup paths in a way that single link failure can be recovered. A and B are label switch routers. Say each link is of unit bandwidth and each LSP request is also of unit bandwidth. L1 and L2 are two active paths. L1b is the backup for L1 and L2b is the backup for L2. L1b and L2b can be accomodated on the same link by sharing the bandwidth. Clearly, if either one of L1 or L2 fail the system can recover.

2.2 Sharing Backup : Single Node Failure Recovery

Figure 2 illustrates a simple case of sharing of backup paths in a way that single node failure can be recovered.

```
               ----            L2            ----
              | A  |-------------------------| B  |
               ----                          ----
                 |                            |
                 |L2b                         |L2b
                 |                            |
                 |                            |
               ----         L1b  L2b         ----
              | C  |-------------------------| D  |
               ----                          ----
                 |                            |
```

```
                   |L1b                            |L1b
                   |                               |
                   |                               |
                  ----     L1     ----     L1     ----
                 | E  |---------| F  |-----------| G  |
                  ----            ----             ----
```

Figure 2 : Sharing backup links with node failure recovery

A,B, ... G are label switch routers. L1 is an active path along 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 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.

2.3 Local Restoration: Single Link/Node Recovery

Local restoration (SONET like recovery constraints), can be achieved by providing intermediate nodes with a backup path. The intermediate nodes can switchover to the backup path immediately on getting a failure indication.

Figure 3 illustrates an example of local restoration for single link failure recovery.

Clearly, from section 2.1, sharing of backup paths can be done in this case to achieve single link/node failure recovery. In fact, a further degree of sharing can be achieved by sharing of links between segments of the backup paths A-D-B and B-E-C (intra demand sharing).

```
                        ----                   ----
                   ____| D  |____            _____| E  |____
               L1b/     ----     \L1b   /L1b   ----     \L1b
                 /                \    /                 \
                ----      L1       ----       L1        ----
               | A  |-------------| B  |---------------| C  |
                ----               ----                 ----
```

Figure 3 : Local restoration of link failure

Similarly Figure 4 illustrates an example of local restoration for single node failure recovery.

```
                          ----                            ----
         ________________| E  |_____________      _____| F  |____
    L1b/                  ----              \L1b   /L1b   ----     \L1b
      /                                      \    /                 \
     ----       L1        ----       L1       ----       L1        ----
    | A  |--------------| B  |---------------| C  |--------------| D  |
     ----                 ----                ----                 ----
                            \                                     /
                             \L1b             ----               /L1b
                              \_______________| G  |____________/
                                              ----
```

Figure 4 : Local restoration of single node/link failure

Clearly, from sections 2.1 and 2.2, Figure 3, and Figure 4 backup sharing (intra and inter demand sharing) can be done in this case as well.

2.4 A Simple Algorithm for Calculating Shared Backup Paths

Terminology: Say for link (i,j)
i) the cumulative bandwidth allocated for active paths is F(i,j)
ii) the cumulative bandwidth allocated for backup paths is G(i,j)
iii) 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

i) 0 if { M+b < G(u,v) } else
ii) b if { G(u,v) <= M and b <= R(u,v) } else
iii) M+b - G(u,v) if { M <= G(u,v) and M+b <= G(u,v)+R(u,v) } else
iv) 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 lack of an active path or a backup path with finite cost represent failure conditions.

2.5 Distributed Determination of Shared Bandwidth on a Link for Backup Paths

In the distributed scenario where path computation for an LSP occurs at its ingress node, sharability of bandwidth on the backup path cannot be determined during path computation. This is because two active paths L1 and L2 can share bandwidth on a link on their backup paths only if L1 and L2 are failure entity (e.g., link or node or SRLG) disjoint. Thus, determination of backup bandwidth sharability requires global information about all LSPs in the network which is not available at the ingress node.

Note that a node in the network can maintain (in a local database) explicit hop and bandwidth information about the LSPs whose active or backup paths pass through it. This information can be disseminated during signaling of the paths through this node, using extensions to signaling protocols like RSVP [10], [13] or CR-LDP [14] to carry the explicit hop information for active and backup paths. Using explicit information about the active paths whose backup paths traverse a link adjacent to a given node, this node can locally determine the amount of shared backup bandwidth that needs to be allocated on all links adjacent to it (as described below).

Consider the case where the failure entity to protect against is a link, i.e., single link failure recovery as discussed in Section 2.1. For a link A-B adjacent to nodes A and B, the set of active paths whose backup paths traverse link A-B are known at both nodes A and B. For each link X-Y that is on any of these active paths, compute the sum of the bandwidths of the LSPs whose active paths fail when this link goes down. The maximum value over all such links X-Y is the amount of shared bandwidth that needs to be reserved on link A-B for backup paths. Note that this computation can be done at either node A or B using its local database which contains information about all LSPs whose active or backup paths traverse link A-B.

When the signaling message for provisioning a new backup path arrives at a node over an adjacent link, the node decides whether to increase the shared bandwidth on this link using the above computation. Note that in this case, the above computation needs to be repeated for only those links X-Y that belong to the active path of this new LSP.

This is because the failure of other links (i.e., those which do not belong to the active path of this new LSP) will affect the active paths of other LSPs whose backup paths traverse this link and this

has been taken into account during previous computations of the required backup bandwidth on this link.

The above framework for distributed determination of shared bandwidth for backup paths can be used together with any path computation algorithm at the ingress node which may or may not use (partial) information available about the bandwidth used on each link for active and backup paths. Local computation of shared bandwidth for backup paths allows optimal allocation of shared backup bandwidth on a link after the active and backup paths for an LSP have been computed at the link-level at the ingress node.

3. Requirements for Shared Backup LSP Restoration

Requirements from link state routing protocols and signaling protocols is briefly described in this section.

Aggregate information about a link that has to be conveyed by a link state routing protocol should consist of
i) The total bandwidth used on the link for active LSPs
ii) Total bandwidth used on the link for backup LSPs
iii) Total available bandwidth on the link

The signaling protocol information elements should consist of
i) The setup information and procedures for a backup LSP
ii) The association between the active and backup LSP
iii) Explicit hop information about the active and backup LSPs at the level of the failure entity to protect against, e.g., link or node.

Every node needs to maintain explicit hop information (at the failure entity level) about all active LSPs whose backup LSPs pass through that node.

4. Security Considerations

This document raises no new security issues.

5. Acknowledgements

The authors would like to thank Vishal Sharma and Roch Guerin for their comments on this work.

6. References

[1] Moy, J,, "OSPF Version 2" RFC 2328, April 1998.

[2] "Intermediate System to Intermediate System Intra-Domain Routeing Exchange Protocol for use in Conjunction with the Protocol for Providing the Connectionless mode Network Service (ISO 8473)", ISO DP 10589, February 1990.

[3] Makam, V., Sharma, V., Huang, C., Owens, K., Mack-Crane, B., et al, "A Framework for MPLS-based Recovery," Work in Progress, Internet Draft <draft-makam-mpls-recovery-frmwrk-00.txt>, February 2000.

[4] Lakshman, T.V., Kodialam, M., "Dynamic Routing of Bandwidth Guaranteed Tunnels with Restoration", Proceedings of INFOCOM 2000, April 2000.

[5] Lakshman, T.V., Kodialam, M., "Dynamic Routing of Bandwidth Guaranteed Tunnels Using Aggregated Link Usage Information", To be published in Proceedings of INFOCOM 2001.

[6] Haskin, D., Krishnan, R., "A Method for Setting an Alternative Label Switched Paths to Handle Fast Reroute", Work in Progress, Internet Draft <draft-haskin-mpls-fast-reroute-04.txt>, May 2000.

[7] Rosen, E., Viswanathan, A., and Callon, R., "Multiprotocol Label Switching Architecture", Work in Progress, Internet Draft <draft-ietf-mpls-arch-06.txt>, August 1999.

[8] Callon, R., Doolan, P., Feldman, N., Fredette, A., Swallow, G., Viswanathan, A., "A Framework for Multiprotocol Label Switching", Work in Progress, Internet Draft <draft-ietf-mpls-framework-05.txt>, September 1999.

[9] Braden, R., Zhang, L., Berson, S., Herzog, S., "Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification", RFC 2205, September 1997.

[10] Awduche, D. et al, "RSVP-TE: Extensions to RSVP for LSP Tunnels", Work in Progress, Internet Draft <draft-ietf-mpls-rsvp-lsp-tunnel-07.txt>, August 1999.

[11] Andersson, L., Doolan, P., Feldman, N., Fredette, A., Thomas, B., "LDP Specification", Work in Progress, Internet Draft <draft-ietf-mpls-ldp-06.txt>, September 1999.

[12] Jamoussi, B., et al, "Constraint-Based LSP Setup using LDP", Work in Progress, Internet Draft <draft-ietf-mpls-cr-ldp-03.txt>, September 1999.

[13] Kini, S., Kodialam, M., Lakshman, T.V., Villamizar, C., "ReSerVation Protocol with Traffic Engineering extensions: Extensions for Label Switched Path Restoration", Work in

Progress, Internet Draft <draft-kini-rsvp-lsp-restoration-00.txt>, April 2001.

[14] Owens, K., et al, "Extensions to CRLDP for MPLS Path Protection", Work in Progress, Internet Draft <draft-owens-crldp-path-protection-ext-00.txt>, November 2000.

7. Author's Addresses

Sriganesh Kini
Mahi Networks
1039 North McDowell Blvd.
Petaluma, CA 94954.
Phone: 707 283 1257
Email: skini@mahinetworks.com

Murali Kodialam
Lucent Technologies, Bell Labs
Room 4D-525, 101 Crawfords Corner Road
Holmdel, NJ 07733-3030
Phone: 732 949 6296
Email: muralik@dnrc.bell-labs.com

T.V.Lakshman
Lucent Technologies, Bell Labs
Room 4D-531, 101 Crawfords Corner Road
Holmdel, NJ 07733-3030
Phone: 732 949 4778
Email: lakshman@dnrc.bell-labs.com

Sudipta Sengupta
Tellium, Inc.
2 Crescent Place PO Box 901
Oceanport, NJ 07757-0901
Phone: 732 483 2837
Email: sudipta@tellium.com

Curtis Villamizar
Avici Systems
Email: curtis@avici.com

Full Copyright Statement