Operation Administration and Maintenance in MPLS based Ethernet Networks

Jordi Perelló, Luis Velasco, Gabriel Junyent
Optical Communication Group - Universitat Politècnica de Cataluya (UPC)
E-mail: jper7775@alu-etsetb.upc.edu ,{luis.velasco, gabriel.junyent}@tsc.upc.edu

In this paper we present a global vision of the Operation, Administration and Maintenance mechanisms for MPLS networks suggested by the ITU-T and the IETF. The organization of the present paper shows three well separated parts: The OAM requirements, a description of these mechanisms and finally the connection between the OAM and the network protection mechanisms.

This paper is intended to provide a general vision of the currently developed or in development OAM mechanisms for MPLS networks and how these mechanisms can help in the implementation of Metropolitan Ethernet Networks (MEN).

1 INTRODUCTION

The Ethernet migration from LAN environments to MAN environments presents some problems due to its LAN nature: end-to-end QoS not provided, lack of fast failure detection, lack of OAM mechanisms and scalability problems.

The introduction of Layer-2 MPLS helps to overcome the mentioned problems. The OAM mechanisms provide fault detection and localization capacity to, once the failure is detected, initiate the network protection in order to provide an end-to-end quality of service.

The remain of this paper exposes the OAM mechanisms purposed by the ITU-T and the IETF.

2 OAM REQUIREMENTS

The standardization organizations ITU-T [1] and IETF [5] studies, and also different documents published by the IEEE [11][12], show the requirements that the OAM mechanisms for MPLS networks may accomplish:

- **LSP connectivity verification capacity**. The LSP connectivity between the ingress and the egress nodes should be tested at any moment.
- **LSP fault detection**: The OAM mechanisms should be capable to detect the following errors: Simple loss of connectivity, LSP swapping, LSP merging and loops (see Figure 1).
- **LSP defect notification**: Once the failure is detected, it should be notified to the client layers for alarm suppression. Moreover, the fault should be notified to the points of repair (POR) to initiate the traffic switching to a backup LSP (see chapter 8. Fault Notification and Network Protection).
- **LSP defect localization**. The network should be able to support automatic fault localization in order to avoid long manual detections. The localization process should be initiated just after fault detection.

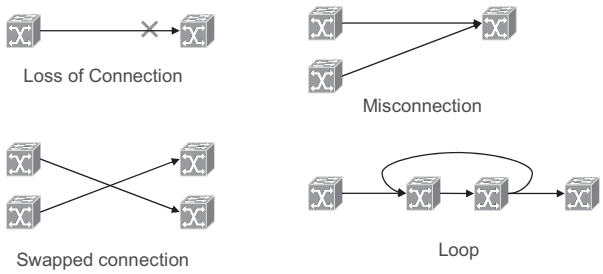

Figure 1 - Types of LSP defect.

- **Client layer alarm suppression due to server layer failures**. An error occurred in a server layer can lead to multiple unnecessary alarms in multiple client layers (Alarm Storming). The OAM mechanisms should be capable to suppress these alarms by notifying the current attendance of the failure,
- **SLA (Service Level Agreement) monitorization**. The SLA monitoring ensures that traffic accomplishes the committed quality of service. Some of the measurable parameters could be:
  - LSP availability
  - Latency
  - Packet Loss
  - Jitter
- **Independence between the OAM control plane and the OAM data plane**. A separation between the control plane and the data plane should be present. The OAM packets should follow the same path as the user's traffic.
- **Frequency configuration capacity**. The different OAM mechanisms should be configurable in terms of frequency of execution to permit the network operator to provide different qualities of service for different services. The frequency of execution is closely related to detection and recovery times.

In addition to the requirements previously cited, we can also mention other additional requirements such as:

- **Scalability**: The OAM mechanisms should be easily scalable to large networks.
- **Reliability**: The OAM packets should be reliable, that is, they should contain error detection and correction mechanisms.
- **Compatibility**: It is important to maintain compatibility with previous versions. The nodes which do not understand the new mechanisms have to omit them.

The next step in this paper will be the description of the different mechanisms supported by the ITU-T and the IETF.

3 CONNECTIVITY VERIFICATION AND FAST FAILURE DETECTION

The mechanisms called *Connectivity Verification* (CV) and *Fast Failure Detection* (FFD) proposed by the ITU-T [2] allow the detection and diagnosis of possible errors occurred in the MPLS layer or lower layers.

The CV packet flow, with a sending frequency of 1 packet per second, has its origin at the ingress node of the LSP and goes to the egress node of that LSP, as we can see in Figure 2. The egress node checks if the incoming CV packets belong to the correct ingress node for that LSP.

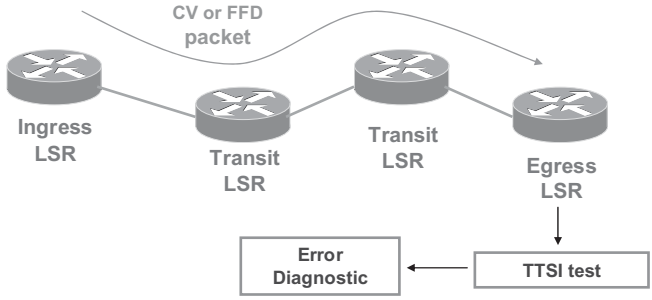

Figure 2 - CV and FFD packets path.

In Figure 3 we can see the CV packet structure. The TTSI (*Trail Termination Source Identifier*) identifies the ingress node of the LSP being tested, in order to perform the diagnosis of the possible reception errors.

The errors that CV mechanism can detect are:

- **dLOCV**: Lack of packets with the expected TTSI.
- **dTTSI_Mismatch**: Observation of packets with a TTSI different than the expected one.
- **dTTSI_Mismerge**: Observation of packets with the expected TTSI and packets with an unexpected TTSI.
- **dExcess**: Increase of the reception packets rate with the expected TTSI.

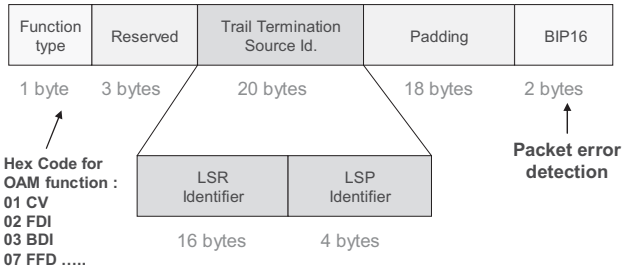

Figure 3 - CV packet structure.

The operation of the FFD mechanism is the same as CV. The main difference, as opposite to CV, is that FFD allows the variation of FFD packets sending frequency, enabling fast failure detection. The recommended value is 20 packets per second (one packet every 50ms). In the FFD frame there is an octet used to choose the sending frequency.

The time of failure detection is 3 seconds for CV and *3x* for FFD, being *x* the period between two FFD packets. Within this detection time, a possible loss of one CV or FFD packet is considered.

4 FORWARD DEFECT INDICATION AND BACKWARD DEFECT INDICATION

The FDI (Forward Defect Indication) and BDI (Backward Defect Indication) mechanisms are purposed by ITU-T [2].

FDI main objective is the alarm suppression in the client LSPs affected by a defect. The FDI packets are sent forward, with a frequency of 1 packet per second, from the first node that detects the failure. If the defect is located at the server layer, it will be the next node from the failure. If the defect is located in the MPLS layer, it will be the next LSP termination at the same level where the defect has occurred.

The BDI mechanism is used to inform the ingress node of the LSP that an error has been detected at destination. The sending frequency of BDI packets is 1 packet per second. BDI requires a return path, which could be a dedicated path for that LSP, a shared path for various LSPs or a non-MPLS return path.

Figure 4 shows the FDI and BDI operation.

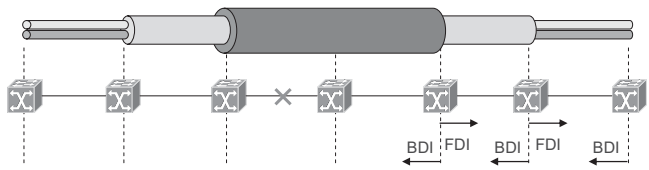

Figure 4 - FDI and BDI mechanisms.

As we can see in Figure 5, the FDI frame includes the error that has been detected (dServer, dPeerME, dLOCV, dTTSI_Mismatch, dTTSI_Mismerge, dExcess, dUnknown, etc.) and the identity of the element where the error has been detected.

Function type	Reserved	Defect Type	TTSI (Optional)	Defect position	Padding	BIP16
1 byte	1 byte	2 bytes	20 bytes	4 bytes	14 bytes	2 bytes

Figure 5 - FDI packet structure.

FDI and BDI can be useful when monitoring network availability or in traffic switching for protection mechanisms (see chapter 8. Fault Notification and Network Protection).

5 MPLS LSP PING

The MPLS LSP Ping proposed by the IETF [6] is intended to verify that packets belonging to a certain FEC (*Forwarding Equivalence Class*) end their MPLS path in an egress node appropriated for that FEC. The *MPLS LSP Ping echo request packets* are sent from the ingress node to the egress node, following the same path as the packets belonging to the FEC being tested (see Figure 6).

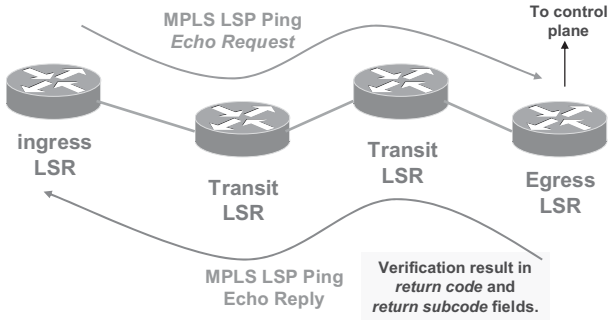

Figure 6 – Example of *MPLS LSP Ping*.

The *echo request* and *echo reply* packets contain a specific field for the inclusion of TLVs (*Type Length and Value*). The TLVs allow the inclusion of variable length information in the packets (see Figure 7). The FEC under test is included in the *echo request* using the *FEC stack TLV*. When using *penultimate hop popping* (PHP) with a label stack containing only one label, the only way to know which FEC is being verified by the egress node is including that FEC in the message.

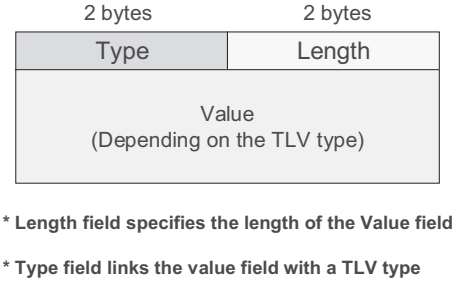

Figure 7 – Example of a TLV structure.

LSP Ping presents two operational modes: *Basic connectivity check* and *Traceroute*. In the former, the *echo request* packet reaches the egress node and then it is sent to the control plane which will verify if that egress node is a valid egress for that FEC. Once the verification is done, the egress node will send a *MPLS echo reply* to the ingress node with the *return code* reflecting the verification conclusion. The *return code* and the *return subcode* fields are used to code the result of the verification and the point of the label stack where this event has been verified. In the *Traceroute* mode, the packet is

sent to the control plane of every transit LSR where a verification will be performed. Every transit LSR will verify that it is a valid transit LSR for that FEC.

The normal procedure when testing LSP connectivity will be the sending of *MPLS echo requests* in *basic connectivity check* mode and, when no receiving any reply or receiving a reply with a *return code* reflecting an erroneous verification, initiating the *traceroute* mode in order to localize the error.

In the *traceroute* mode (see Figure 8), the ingress node (in this example, LSR1) sends packets incrementing by 1 the TTL field (1,2,3,4,…) searching the expiration of that field when the packet will be received by the node that has to carry out the verification. In the example, the verification that LSR2 is a valid transit LSR for the FEC under test is done firstly and then is verified that LSR3 is an egress node for that FEC.

In the *traceroute* mode, the TLV called *Downstream Mapping* is used. When a LSR carries out a verification, for example LSR2 verifies that is a valid transit LSR for that FEC, it includes *downstream mapping* objects to identify each interface where that FEC could be sent. When the LSR is the egress LSR for that FEC, it will not include any *downstream mapping* object.

When receiving the *echo reply*, the ingress LSR will send another echo request incrementing by 1 the TTL field and copying the *downstream mapping* from the *echo reply* received. In that way we can verify all the different paths that the given FEC can follow.

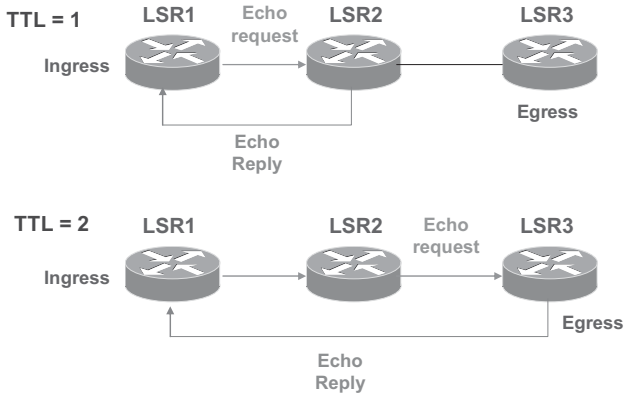

Figure 8 – Example of *Traceroute* mode.

6 BIDIRECTIONAL FORWARDING DETECTION

As we have seen, LSP Ping can detect data plane failures and then carry out an analysis of the data plane against the control plane. In contrast, BFD (*Bidirectional Forwarding Detection*), proposed by the IETF [7][8], is only intended to detect data plane defects incurring in a lower computational cost, providing fast failure detection (<1s in contrast with the various seconds of *LSP Ping*). It also gives the opportunity to implement failure detection for a larger number of LSPs. Moreover, due to its fixed packet size is easier to implement in hardware than *LSP Ping*.

Using BFD together with *LSP Ping* can help us to obtain the benefits of both. *LSP Ping* is used to initiate the BFD session, while BFD is used for failure detection. Periodically *LSP Ping* can be used to verify the data plane against the control plane.

BFD can be used in three different modes:

- **Asynchronous mode**: Packets are sent between the ingress and the egress LSR. When a determined number of packets are lost, the session is considered to be down.
- **Demand mode**: In demand mode is assumed that the two LSRs have an alternative way to check the connectivity between them. Once the session is established, they stop sending packets between them. When one of them wishes to verify the LSP connectivity, it sends various packets to the other and then the protocol turns to active.
- **Echo mode**: In echo mode one of the endings (the ingress or the egress LSR) sends BFD packets through the way that the packets are looped through the remote LSR. When a number of packets are lost, the session is considered to be down. The echo mode can run asynchronously or under demand.

If a LSP is associated to multiple FECs, every FEC has its own session. In BFD packets (see Figure 9) some fields (*Discriminators*) are included to associate BFD packets with the correct session.

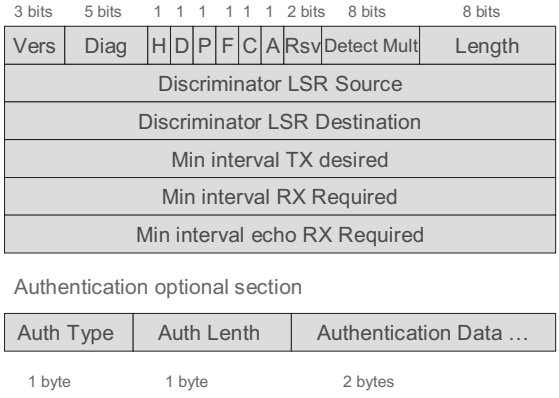

Figure 9 - BFD packet structure.

The different transmission intervals (that determine the detection time) of the BFD packets are negotiated continuously, so that they can be varied at any moment. If the LSP is bidirectional, the negotiation is independent on both directions. The system with the lower transmission/reception rate will determine the transmission/reception rate of the BFD packets.

7 LABEL SWITCHING ROUTER SELF TEST

The LSR Self Test proposed by the IETF [9] defines a mechanism that gives a LSR the label association testing capacity and also the connectivity verification capacity between that LSR and its upstream and downstream LSRs.

A new FEC type called *Loopback FEC* is defined. It permits the upstream node to advertise the loopback label that will be used in the test. This label allows the upstream node to loop the test packets without any control plane intervention. The packets will be looped with only looking the MPLS label.

The *LSP Ping echo request and LSP Ping echo reply* messages are extended for its use in the LSR Self Test. This extension leads to *MPLS Data Plane Verification Request* (MPLS DPVRq) and *Reply* (MPLS DPVRp).

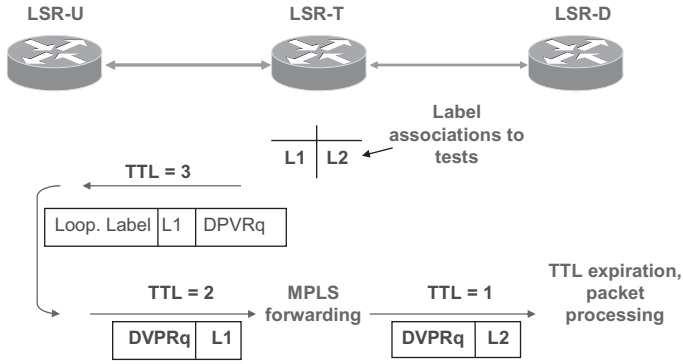

Figure 10 – Example of *DPVRq* in a *LSR Self Test*.

An example of the MPLS *DPVRq* route is shown in Figure 10. Three LSRs are involved in the test: The LSR that initiates the test (LSR-T) and its upstream (LSR-U) and downstream (LSR-D) LSRs.

When initiating the test, LSR-U announces to LSR-T the *loopback label* that will be used in that test.

Then, LSR-T creates a *DPVRq* message including the label stack under test and the *loopback label* previously advertised by LSR-U. In the example, LSR-T wants to test the L1 label.

The *loopback label* TTL field is set to expiration when the packet arrives to LSR-D (usually set to 3).

When the packet arrives at LSR-U, the *loopback label* is extracted and its TTL field is copied into the top label of the label stack under test. Then the packet is looped through the interface that the user's packets would follow. In the example, the outgoing packets with the L1 label.

Straight on, when arriving at LSR-T, it does the packet forwarding normally. In the example LSR-T performs normal label swapping from L1 to L2. If all the previous steps have been successful, the *DPVRq* arrives at LSR-D and TTL field is expired. LSR-D saves the interface where the packet have arrived and its label stack.

The test results are sent to LSR-T using a *DPVRp*. This mechanism permits to test neighbour connectivity and correct packet forwarding in LSR-T.

8 FAULT NOTIFICATION AND NETWORK PROTECTION

The relationship between OAM mechanisms and network protection is achieved through fault notification. Once the failure is detected, the MPLS network should be able to switch user's traffic in order to provide the committed quality of service.

In a MPLS network we can have different protection schemes [3][10]:

- **1+1 protection**: In this case, two LSPs are set between the ingress LSR (*Path Source LSR, PSL*) and the egress LSR(*Path Merge LSR, PML*). The traffic is sent simultaneously through the two LSRs and the PML chooses from which LSP wants to receive the traffic.
- **1:n protection**: In this case, two LSPs between PSL and PML are defined too, but PSL only sends user's traffic through the main LSP. When no failure is detected, the secondary LSP can be used to carry *best-effort* traffic. When a failure is detected, PSL stops forwarding *best-effort* traffic and switches the main traffic to the secondary LSP.

The node detecting the failure, if no direct action can be done by itself, should be capable of notifying that failure to the *point of repair* (POR), where the switching has to be done. Depending on what kind of protection is being used, POR can be the PSL LSR (1:n) or PML LSR (1+1).

The IETF [10] suggests using a FIS (*Failure Indication Signal*) in fault notification (see Figure 11). The FIS can be transmitted through the data plane or the control plane. Some transmitting options could be:

- Through the control plane, using the hello messages sent between neighbour LSRs with an additional field used for fault notification until reaching the POR. One example could be the hello messages sent by the RSVP protocol between adjacent nodes.
- Through the data plane using any specific packet for fault notification. This packet could be sent using the same LSP if bidirectional or using any specific return path.

If the network uses pre-signaled backup paths, the crucial time intervals to perform the network recovery are the *detection time* (interval of time since the failure occurres and its detection) and the *notification time* (interval of time compressed between the sending start of notification packets in the detecting node and the reception of them in the PSL) [13][14].

If using OAM mechanisms for failure detection, *detection time* depends on the used mechanism. On the other hand, notification involves the necessary *transmission time* to cross all the links until the PSL is reached, and the *queuing time* (the time the packets have to wait in the queues until they are sent to the output interface) in all intermediate nodes. While notification can be performed using RSVP packets, this option seems not to be optimal. RSVP packets do not necessary follow the shortest reverse path to the PSL, so *transmission time* is not optimized. In addition to, RSVP packets are highly affected by queuing times.

In contrast with RSVP notification, IETF [14] purposes *flooding techniques*. Once the failure is detected, notification packets are flooded through the control plane.

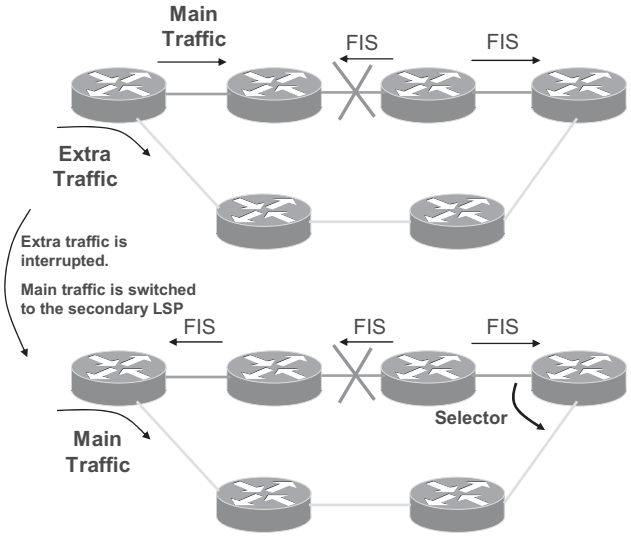

Figure 11 – Example of FIS notification in a 1:1 protection.

Using priority queuing, the *queuing time* of the flooded packets can be 0 and, if using a shortest path algorithm to calculate the reverse path, the *transmission time* would be minimal. Moreover, since all nodes are informed about the network failure, they can use this information when calculating routes.

The ITU-T [2] suggests using the FDI and BDI mechanisms for fault notification. FDI allows performing downstream fault notification (until PML is reached). On the other hand, BDI can be used for upstream fault notification (when PSL has to initiate the protection switching).

Notification and recovery times are also hardly affected on the recovery model being used. Due to the length limitations of this paper, the different recovery models are out of the scope of the present. For further information read [13].

9 CONCLUSIONS

The present paper shows the benefits that MPLS OAM mechanisms can provide to Ethernet MPLS-based networks. The choice between one or another will depend on the requirements of the services being carried by the network.

CV and MPLS LSP Ping seem not to be adequate to support fast failure detection due to its slow packet transmission rates. FFD and BFD are more appropriate due to its potentially faster transmission rate. MPLS LSP Ping packets can be sent periodically to test the control plane, but giving the fault detection task to a faster protocol.

In the notification, as explained in the previous chapter, flooding techniques seem to be the most well situated to be implemented due to its lower notification time.

Finally a summary where the requirements (described in chapter 2.- OAM Requirements) accomplished by the studied mechanisms is shown (see Figure 12).

	CV	FFD	FDI BDI	LSP Ping Traceroute	BFD	LSR Self Test*
LSP connectivity verification	√	√		√	√	√
LSP fault detection	√	√		√	√	√
LSP defect notification			√	√		
LSP defect localization			√	√		
Client layer alarm suppression			√			
SLA measurement capacity						
Control plane vs. Data plane independence.	√	√	√	√	√	√
Frequency config. capacity		√			√	

*** LSR Self Test do not verify LSPs, but connectivity between LSR-T and LSR-U, LSR-D.**

Figure 12 - Summary.

10 REFERENCES

[1] ITU-T Y.1710, “Requirements for OAM functionality for MPLS networks”, Nov. 2002.

[2] ITU-T Y.1711, “OAM mechanism for MPLS networks”, Feb. 2004

[3] ITU-T Y.1720, “Protection Switching for MPLS Networks”, Sept. 2003.

[4] ITU-T Y.1730, “Requirements for OAM functions in Ethernet-based networks and Ethernet services”, Jan. 2004.

[5] T. D. Nadeau *et al.*, “OAM Requirements for MPLS Networks,” IETF Internet draft draft-ietf-mpls-oam-requirements-05.txt, Dec 2004.

[6] Kompella *et al.*, “Detecting MPLS Data Plane Failures” IETF Internet draft draft-ietf-mpls-lsp-ping-07.txt, Oct.2004

[7] R. Aggarwal *et al.*, “BFD for MPLS LSPs,” IETF Internet draft draft-ietf-bfd-mpls-00.txt, Jul. 2004.

[8] D.Katz, D.Ward, “Bidirectional Forwarding Detection” IETF Internet draft draft-ietf-bfd-base-00.txt, Jul. 2004.

[9] G. Swallow, K. Kompella, D. Tappan, “Label Switching Router Self-Test” IETF Internet draft draft-ietf-mpls-lsrself-test-03.txt, Oct. 2004.

[10] V.Sharma, F.Hellstrand, “Framework for Multi-Protocol Label Switching (MPLS)-based Recovery” IETF RFC 3469, Feb. 2003.

[11] D.Cavendish, H.Ohta, H.Rakotoranto “Operation, Administration, and Maintenance in MPLS Networks” IEEE Communications Magazine, Oct. 2004.

[12] Zhi-Wei Lin, “Network OAM Requirements for the New York City Transit Network” IEEE Communications Magazine, Oct. 2004.

[13] J.L. Marzo et al, “Failure Recovery Time Minimization in GMPLS-Based Networks using Segment Protection”.

[14] Richard Rabbat et al, “Fault Notification Protocol for GMPLS-Based Recovery”, IETF Internet draft draft-rabbat-fault-notification-protocol-02.txt, Feb 2003.