Network Working Group
Internet Draft
Expires: May 2005

Thomas D. Nadeau
Monique Morrow
George Swallow
Cisco Systems, Inc.

David Allan
Nortel Networks

Satoru Matsushima
Japan Telecom

December 2004

OAM Requirements for MPLS Networks
draft-ietf-mpls-oam-requirements-05.txt

Status of this Memo

Abstract

As transport of diverse traffic types such as voice, frame

relay, and ATM over MPLS become more common, the ability to detect, handle and diagnose control and data plane defects becomes critical.

Detection and specification of how to handle those defects is not only important because such defects may not only affect the fundamental operation of an Multi-Protocol Label Switching (MPLS) network, but also because they MAY impact service level specification commitments for customers of that network.

This document describes requirements for user and data plane operations and management for MPLS. These requirements have been gathered from network operators who have extensive experience deploying MPLS networks, similarly some of these requirements have appeared in other documents. This draft specifies Operations and Management requirements for Multi-Protocol Label Switching, as well as for applications of Multi-Protocol Label Switching such as pseudowire voice and virtual private network services. Those interested in specific issues relating to instrumenting MPLS for Operations and Management purposes are directed to the Multi-Protocol Label Switching Architecture specification.

1. Introduction

This document describes requirements for user and data plane operations and management (OAM) for Multi-Protocol Label Switching (MPLS). These requirements have been gathered from network operators who have extensive experience deploying MPLS networks. This draft specifies OAM requirements for MPLS, as well as for applications of MPLS.

No specific mechanisms are proposed to address these requirements at this time. The goal of this draft is to identify a commonly applicable set of requirements for MPLS OAM at this time. Specifically, a set of requirements that apply to the most common set of MPLS networks deployed by service provider organizations today. These requirements can then be used as a base for network management tool development and to guide the evolution of currently specified tools, as well as the specification of OAM functions that are intrinsic to protocols used in MPLS networks.

Comments should be made directly to the MPLS mailing list at mpls@lists.ietf.org.

2. Document Conventions

2.1 Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119.

Defect: Any error condition that prevents an LSP functioning correctly. For example, loss of an IGP path will most likely also result in an LSP not being able to deliver traffic to its destination. Another example is the breakage of a TE tunnel. These MAY be due to physical circuit failures or failure of switching nodes to operate as expected.

Multi-vendor/multi-provider network operation typically requires agreed upon definitions of defects (when it is broken and when it is not) such that both recovery procedures and service level specification impacts can be specified.

Head-end Label Switch Router (LSR): The beginning of a label switched path.

Probe-based-detection: Active measurement using a tool such as LSP ping.

Collecting traffic: Passive measurement of network traffic.

Head-end Label Switching Router (LSR): The beginning of a label switched path. A head-end LSR is also referred to as an Ingress Label Switching Router.

Probe-based-detection: Active measurement using a tool such as LSP ping.

Collecting traffic: Passive measurement of network traffic.

propagation latency: delay added by the propagation of the packet through the link (fixed value that depends on the distance of the link and the propagation speed).

transmission latency: delay added by the transmission of the packey over the link i.e. the time it takes put the packet over the media (value that depends of

the link throughput and packet length).

processing latency: delay added by all the operations related to the switching of labeled packet (value is node implementation specific and may are considered as fixed and constant for a given equipment).

queuing/buffering latency: delay caused by packet queuing (value is variable since depending on the packet arrival rate in addition to the dependance on the packet length and the link throughput).

node latency: delay added by the network element resulting from of the sum of the transmission, processing and queuing/ buffering latency

one-hop delay: fixed delay experienced by a packet to reach the next hop reesulting from the of the propagation latency, the transmission latency and the processing latency.

minimum path latency: sum of the one-hop delays experienced by the packet when travelling from the ingress to the egress LSR.

variable path latency (jitter): sum of the delays caused by the queuing latency experienced by the packet at each node over the path.

2.2 Acronyms

CE: Customer Edge

SP: Service Provider

ECMP: Equal Cost Multipath

LSP: Label Switch Path

LSR: Label Switch Router

OAM: Operations and Management

RSVP: Resource reSerVation Protocol

LDP: Label Distribution Protocol

DoS: Denial of service

3. Motivations

MPLS OAM has been tackled in numerous Internet drafts. However as of this writing, existing drafts focus on single provider solutions or focus on a single aspect of the MPLS architecture or application of MPLS. For example, the use of RSVP or LDP signaling and defects MAY be covered in some deployments, and a corresponding SNMP MIB module exists to manage this application; however, the handling of defects and specification of which types of defects are interesting to operational networks MAY not have been created in concert with those for other applications of MPLS such as L3 VPN. This leads to inconsistent and inefficient applicability across the MPLS architecture, and/or requires significant modifications to operational procedures and systems in order to provide consistent and useful OAM functionality which do not create inconsistencies with existing solutions. As MPLS has matured, relationships between providers has become more complex. Furthermore, the deployment of multiple concurrent applications of MPLS is common place, leading to a need to consider broader and more uniform solutions, rather than very specific ad hoc point solutions.

4. Requirements

The following sections enumerate the OAM requirements gathered from service providers who have deployed MPLS and services based on MPLS networks. Each requirement is specified in detail to clarify its applicability. Although the requirements specified herein are defined by the IETF, they have been harmonized with requirements gathered by other standards bodies such as the ITU [Y1710].

4.1 Detection of Label Switch Path Defects

The ability to detect defects in a broken Label Switch Path (LSP) SHOULD not require manual hop-by-hop troubleshooting of each LSR used to switch traffic for that LSP. For example, it is not desirable to manually visit each LSR along the data plane path used to transport an LSP; instead,this function SHOULD be automated and able to be performed at some operator specified frequency from the origination point of that LSP. This implies solutions that are interoperable as to allow for such automatic operation. Furthermore, the automation of path liveliness is desired in cases where large numbers of LSPs might be tested. For example, automated ingress LSR to egress LSR testing functionality is desired for some LSPs. The goal is to detect LSP path defects before customers do, and this requires detection of

LSP defects in a "reasonable" amount of time. One useful definition of reasonable is both predictable and consistent.

Synchronization of detection time bounds by tools used to detect broken LSPs is required. Failure to specifying defect detection time bounds may result in an ambiguity in test results. If the time to detect is known, then automated responses can be specified both with respect to and with regard to resiliency and service level specification reporting. Further, if synchronization of detection time bounds is possible, an operational framework can be established that can guide the design and specification of MPLS applications.

Although ICMP-based ping [RFC792] can be sent through an LSP, the use of this tool to verify the defect free operation of an LSP has the potential for returning erroneous results (both positive and negative). For example, failures may occur when inconsistencies exist within the IP or MPLS forwarding tables, in the MPLS control and data planes or LSP. Failures may also result from defects with the reply path (i.e., a reverse path does not exist) used to return a response to a test message. As an example of a false positive, consider the case where the MPLS data plane flows through a network node using a different output line card than the data plane uses to reach the next-hop neighbor. Also assume that although the control plane is functional, the data plane on the output line card where data traffic is programmed to exit the device is defective. Now, if an LSP is signaled using this node, any test based solely on the control plane's view of the world (i.e., ICMP-based) will return with a false positive result because although the control plane traffic at the node in the example would be forwarded correctly, the actual data plane switching at the node in the example would misroute or drop any traffic transmitted onto that LSP. An example of a false negative case would be when a functioning return path does not exist. In this case, neither a positive nor a negative reply will be received by the sender. Therefore any detection mechanisms that depend on receiving status via a return path SHOULD provide multiple return options with the expectation that one of them will not be impacted by the original defect.

The OAM packet MUST follow exactly the customer data path in order to reflect path liveliness used by customer data. Particular cases of interest are forwarding mechanisms such as equal cost multipath (ECMP) scenarios within the operator's network whereby flows are load-shared across parallel (i.e., equal IGP cost) paths. Where the customer traffic MAY be spread over multiple paths, what is required is to be able to detect failures on any of the path permutations. Where the spreading mechanism is payload specific, payloads need to have forwarding that is common with the traffic under test. Satisfying these requirements introduces complexity

into ensuring that ECMP connectivity permutations are exercised, and that defect detection occurs in a reasonable amount of time.

4.2 Diagnosis of a Broken Label Switch Path

The ability to diagnose a broken LSP and to isolate the failed component (i.e., link or node) in the path is required. For example, note that specifying recovery actions for misbranching defects in an LDP network is a particularly difficult case. Diagnosis of defects and isolation of the failed component is best accomplished via a path trace function which can return the the entire list of LSRs and links used by a certain LSP (or at least the set of LSRs/links up to the location of the defect) is required. The tracing capability SHOULD include the ability to trace recursive paths, such as when nested LSPs are used. This path trace function MUST also be capable of diagnosing LSP mis-merging by permitting comparison of expected vs. actual forwarding behavior at any LSR in the path. The path trace capability SHOULD be capable of being executed from both the head-end Label Switch Router (LSR) and MAY permit downstream path components to be traced from an intermediate mid-point LSR. Additionally, the path trace function MUST have the ability to support equal cost multipath scenarios described above in section 4.1.

4.3 Path characterization

The path characterization function is the ability to reveal details of LSR forwarding operations. These details can then be compared later during subsequent testing relevant to OAM functionality. This would include but is not limited to:

- consistent use of pipe or uniform time to live (TTL) models by an LSR [RFC3443].
- sufficient details that allow the test origin to excersize all path permutations related to load spreading (e.g. ECMP).
- stack operations performed by the LSR, such as pushes, pops, and TTL propagation at penultimate hop LSRs.

4.4 Service Level Agreement Measurement

Mechanisms are required to measure the diverse aspects of Service Level Agreements which include:

- defect free forwarding. The service is considered to be available and the other aspects of performance measurement listed below have meaning, or the service is unavailable and other aspects of performance measurement do not.
- latency - amount of time required for traffic to transit the network

- packet loss
- jitter - measurement of latency variation

Such measurements can be made independently of the user traffic or via a hybrid of user traffic measurement and OAM probing.

At least one mechanism is required to measure the number of OAM packets. In addition, the ability to measure the qualitative aspects of LSPs such as jitter, delay, latency and loss MUST be available in order to determine whether or not the traffic for a specific LSP are traveling within the operator-specified tollerances.

Any method considered SHOULD be capable of measuring the latency of an LSP with minimal impact on network resources. See section 2.1 for definitions of the various qualitative aspects of LSPs.

4.5 Frequency of OAM Execution

The operator MUST have the flexibility to configure OAM parameters insofaras to meet their specific operational requirements.

This includes the frequency of the execution of any OAM functions. The capability to synchronize OAM operations is required as to to permit consistent measurement of service level agreements. To elaborate, there are defect conditions such as misbranching or misdirection of traffic for which probe-based detection mechanisms that incur significant mismatches in the probe rate MAY result in flapping. This can be addressed either by synchronizing the rate or having the probes self-identify their probe rate.

One observation would be that wide-spread deployment of MPLS, common implementation of monitoring tools and the need for inter-carrier synchronization of defect and service level specification handling will drive specification of OAM parameters to commonly agreed on values and such values will have to be harmonized with the surrounding technologies (e.g. SONET/SDH, ATM etc.) in order to be useful. This will become particularly important as networks scale and misconfiguration can result in churn, alarm flapping etc.

4.6 Alarm Suppression, Aggregation and Layer Coordination

Network elements MUST provide alarm suppression functionality that prevents the generation of superfluous generation of alarms by simply discarding them (or not generating them in the first place), or by aggregating them together, and thereby greatly reducing the

number of notifications emitted. When viewed in conjuction with requirement 4.7 below, this typically requires fault notification

to the LSP egress that
MAY have specific time constraints if the application using the LSP independently implements path continuity testing (for example ATM I.610 Continuity check (CC)[I610]). These constraints apply to LSPs that are monitored. The nature of MPLS applications allows for the possibility to have multiple MPLS applications attempt to respond to defects simultaneously. For example, layer-3 MPLS VPNs that utilize Traffic Engineered tunnels, where a failure occurs on the LSP carrying the Traffic Engineered tunnel. This failure would affect he VPN traffic that uses the tunnel's LSP. Mechanisms are required to coordinate network response to defects.

4.7 Support for OAM Interworking for Fault Notification

An LSR supporting the interworking of one or more networking technologies over MPLS MUST be able to translate an MPLS defect into the native technology's error condition. For example, errors occurring over a MPLS transport LSP that supports an emulated ATM VC MUST translate errors into native ATM OAM Alarm Indication Signal (AIS) cells at the termination points of the LSP. The mechanism SHOULD consider possible bounded detection time parameters, e.g., a "hold off" function before reacting as to synchronize with the OAM functions.
One goal would be alarm suppression by the upper layer using the LSP. As observed in section 4.5, this requires that MPLS perform detection in a bounded timeframe in order to initiate alarm suppression prior to the upper layer independently detecting the defect.

4.8 Error Detection and Recovery.

Recovery from a fault by a network element can be facilitated by MPLS OAM procedures. These procesures will detect a broader range of defects than that of simple link and node failures.
Since MPLS LSPs may span multiple routing areas and service provider domains, fault recovery and error detection should be possible in these configuration as well as in the more simplifed single-area/domain configurations.

Recovery from faults SHOULD be automatic. It is a requirement that faults SHOULD be detected (and possibly corrected) by the network operator prior to customers of the service in question detecting them.

4.9 Standard Management Interfaces

The wide-spread deployment of MPLS requires common information modeling of management and control of OAM functionality. This is

reflected in the the integration of standard MPLS-related MIBs (e.g. [RFC3813][RFC3812][RFC3814]) for fault, statistics and configuration management. These standard interfaces provide operators with common programmatic interface access to operations and management functions and their status.

4.10 Detection of Denial of Service Attacks

The ability to detect denial of service (DoS) attacks against the data or control planes MUST be part of any security management related to MPLS OAM tools or techniques.

4.11 Per-LSP Accounting Requirements

In an MPLS network, service providers (SPs) can measure traffic from an LSR to the egress of the network using some MPLS related MIBs, for example. This means that it is a reasonable to know how much traffic is traveling from where to where (i.e., a traffic matrix) by analyzing the flow of traffic. Therefore, traffic accounting in an MPLS network can be summarized as the following three items.

(1) Collecting information to design network

Providers and their customers MAY need to verify high-level service level specifications, either to continuously optimize their networks, or to offer guaranteed bandwidth services. Therefore, traffic accounting to monitor MPLS applications is required.

(2) Providing a Service Level Specification

For the purpose of optimized network design, a service provider may offer the traffic informationr. Optimizing network design needs this information.

(3) Inter-AS environment

Service providers that offer inter-AS services require accounting of those services.

These three motivations need to satisfy the following.

- In (1) and (2), collection of information on a per-LSP basis is a minimum level of granularity of collecting accounting information at both of ingress and egress of an LSP.

- In (3), SP's ASBR carry out interconnection functions as an intermediate LSR. Therefore, identifying a pair of ingress

and egress LSRs using each LSP is needed to determine the cost of the service that a customer is using.

4.11.1 Requirements

Accounting on a per-LSP basis encompasses the following set of functions:

(1) At an ingress LSR accounting of traffic through LSPs beginning at each egress in question.

(2) At an intermediate LSR, accounting of traffic through LSPs for each pair of ingress to egress.

(3) At egress LSR, accounting of traffic through LSPs for each ingress.

(4) All LSRs that contain LSPs that are being measuremented need to have a common key to distinguish each LSP. The key MUST be unique to each LSP, and its mapping to LSP SHOULD be provided from whether manual or automatic configuration.

In the case of non-merged LSPs, this can be achieved by simply reading traffic counters for the label stack associated with the LSP at any LSR along its path. However, in order to measure merged LSPs, an LSR MUST have a means to distinguish the source of each flow so as to disambiguate the statistics.

4.11.2 Scalability

It is not realistic to perform the just described operations by LSRs in a network on all LSPs that exist in a network. At a minimum, per-LSP based accounting SHOULD be performed on the edges of the network -- at the edges of both LSPs and the MPLS domain.

5. Security Considerations

Provisions to any of the tools designed to satisfy the requirements described herin are required to prevent their unauthorized use. Likewise, these tools MUST provide a means by which an operator can prevent denial of service attacks if those tools are used in such an attack.

LSP mis-merging has security implications beyond that of simply being a network defect. LSP mis-merging can happen due to a number of potential sources of failure, some of which (due to MPLS label

stacking) are new to MPLS.

The performance of diagnostic functions and path characterization involve extracting a significant amount of information about network construction which the network operator MAY consider private.

6. IANA Considerations

This document creates no new requirements on IANA namespaces [RFC2434].

7. References

7.1 Informative References

[RFC3812] Srinivasan, C., Viswanathan, A. and T. Nadeau, "MPLS Traffic Engineering Management Information Base Using SMIv2", RFC3812, June 2004.

[RFC3813] Srinivasan, C., Viswanathan, A. and T. Nadeau, "MPLS Label Switch Router Management Information Base Using SMIv2", RFC3813, June 2004.

[RFC3814] Nadeau, T., Srinivasan, C., and A. Viswanathan, "Multiprotocol Label Switching (MPLS) FEC-To-NHLFE (FTN) Management Information Base", RFC3814, June 2004.

[Y1710] ITU-T Recommendation Y.1710, "Requirements for OAM Functionality In MPLS Networks"

[I610] ITU-T Recommendation I.610, "B-ISDN operations and maintenance principles and functions", February 1999

[RFC2434] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations section in RFCs", BCP 26, RFC 2434, October 1998.

[RFC792] Postel, J., "Internet Control Message Protocol", RFC792, September 1981.

[RFC3443] Agarwal, P, Akyol, B., "Time To Live (TTL) Processing in Multi-Protocol Label Switching (MPLS) Networks.", RFC3443, January 2003.

8. Authors' Addresses

Cisco Systems, Inc.
300 Beaver Brook Road
Boxboro, MA 01719
Phone: +1-978-936-1470
Email: tnadeau@cisco.com

Monique Jeanne Morrow
Cisco Systems, Inc.
Glatt-Com, 2nd Floor
CH-8301
Switzerland
Voice: (0)1 878-9412
Email: mmorrow@cisco.com

George Swallow
Cisco Systems, Inc.
300 Beaver Brook Road
Boxboro, MA 01719
Voice: +1-978-936-1398
Email: swallow@cisco.com

David Allan
Nortel Networks
3500 Carling Ave.
Ottawa, Ontario, CANADA
Voice: 1-613-763-6362
Email: dallan@nortelnetworks.com

Satoru Matsushima
Japan Telecom
4-7-1, Hatchobori, Chuo-ku
Tokyo, 104-8508 Japan
Phone: +81-3-5540-8214
Email: satoru@ft.solteria.net

9. Copyright Notice

10. Intellectual Property Statement

on the procedures with respect to rights in RFC documents can be found in BCP 78 and BCP 79.

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at http://www.ietf.org/ipr.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.

11. Disclaimer of Validity

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

12. Copyright Statement

13. Acknowledgment

Funding for the RFC Editor function is currently provided by the Internet Society.

The authors wish to acknowledge and thank the following individuals for their valuable comments to this document:
Adrian Smith, British Telecom; Chou Lan Pok, SBC; Mr. Ikejiri, NTT Communications and Mr.Kumaki of KDDI.
Hari Rakotoranto, Miya Khono, Cisco Systems; Luyuan Fang, AT&T; Danny McPherson, TCB; Dr.Ken Nagami, Ikuo Nakagawa, Intec Netcore, and David Meyer.