

MPLS  
Internet-Draft  
Intended status: Informational  
Expires: September 27, 2015

S. Bryant  
C. Pignataro  
Cisco Systems  
M. Chen  
Z. Li  
Huawei  
G. Mirsky  
Ericsson  
March 26, 2015

MPLS Flow Identification  
draft-bryant-mpls-flow-ident-01

## Abstract

This memo discusses the desired capabilities for MPLS flow identification. The key application that needs this is in-band performance monitoring of user data packets.

## Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of [BCP 78](#) and [BCP 79](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on September 27, 2015.

## Copyright Notice

Copyright (c) 2015 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must

Internet-Draft

MPLS FI

March 2015

include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

## Table of Contents

<a href="#">1.</a>	Introduction . . . . .	<a href="#">2</a>
<a href="#">2.</a>	Requirements Language . . . . .	<a href="#">3</a>
<a href="#">3.</a>	Loss Measurement Considerations . . . . .	<a href="#">3</a>
<a href="#">4.</a>	Delay Measurement Considerations . . . . .	<a href="#">4</a>
<a href="#">5.</a>	Units of identification . . . . .	<a href="#">4</a>
<a href="#">6.</a>	Types of LSP . . . . .	<a href="#">5</a>
<a href="#">7.</a>	Network Scope . . . . .	<a href="#">6</a>
<a href="#">8.</a>	Backwards Compatibility . . . . .	<a href="#">7</a>
<a href="#">9.</a>	Dataplane . . . . .	<a href="#">7</a>
<a href="#">10.</a>	Control Plane . . . . .	<a href="#">8</a>
<a href="#">11.</a>	Manageability Considerations . . . . .	<a href="#">8</a>
<a href="#">12.</a>	Privacy Considerations . . . . .	<a href="#">9</a>
<a href="#">13.</a>	Security Considerations . . . . .	<a href="#">9</a>
<a href="#">14.</a>	IANA Considerations . . . . .	<a href="#">9</a>
<a href="#">15.</a>	Acknowledgements . . . . .	<a href="#">9</a>
<a href="#">16.</a>	References . . . . .	<a href="#">9</a>
<a href="#">16.1.</a>	Normative References . . . . .	<a href="#">9</a>
<a href="#">16.2.</a>	Informative References . . . . .	<a href="#">9</a>
	Authors' Addresses . . . . .	<a href="#">10</a>

## [1.](#) Introduction

This memo discusses the desired capabilities for MPLS flow identification. The key application that needs this is in-band performance monitoring of user data packets.

There is a need to identify flows in MPLS networks for applications such as packet loss and packet delay measurement. A method of loss and delay measurement in MPLS networks was defined in [[RFC6374](#)]. When used to measure packet loss [[RFC6374](#)] depends on the use of the injected OAM packets are used to designate the beginning and the end of the packet group over which packet loss is being measured. Where the misordering of packets from one group relative to the following group, or misordering of one of the packets being counted relative to the [[RFC6374](#)] packet occurs, then an error will occur in the packet loss measurement. In addition, this packet performance system needs to be extended to deal with different granularities of flow and to

address a number of the multi-point cases in which a number of ingress LSRs could send to one or more destinations.

Improvements in link and transmission technologies mean that it may be difficult to assess packet loss using active performance

measurement methods with synthetic traffic, due to the very low loss rate in normal operation. That together with more demanding service level requirements mean that network operators need to be able to measure the loss of the actual user data traffic by using passive performance measurement methods. Any technique deployed needs to be transparent to the end user, and it needs to be assumed that they will not take any active part in the measurement process. Indeed it is important that any flow identification technique be invisible to them and that no remnant of the identification of measurement process leak into their network.

Additionally where there are multiple traffic sources, such as in multi-point to point and multi-point to multi-point network environments there needs to be a method whereby the sink can distinguish between packets from the various sources, that is to say, that a multi-point to multi-point measurement model needs to be developed.

## 2. Requirements Language

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 [[RFC2119](#)].

## 3. Loss Measurement Considerations

Modern networks, if not oversubscribed, normally drop very few packets, thus packet loss measurement is highly sensitive to counter errors. Without some form of coloring or batch marking such as that proposed in [[I-D.tempia-ippm-p3m](#)] it may not be possible to achieve the required accuracy in the loss measurement of customer data traffic. Where accuracy better than the data link loss performance of a modern optical network is required, it may be economically advantageous, or even a technical requirement, to include temporal marking.

Where this level of accuracy is required and the traffic between a source-destination pair is subject to ECMP a demarcation mechanism is needed to group the packets into batches. Once a batch is correlated at both ingress and egress, the packet accounting mechanism is then able to operate on the batch of packets which can be accounted for at both the packet ingress and the packet egress. Errors in the accounting are particularly acute in LSPs subjected to ECMP because the network transit time will be different for the various ECMP paths since:

- a. The packets may traverse different sets of LSRs.

- b. The packets may depart from different interfaces on different line cards on LSRs
- c. The packets may arrive at different interfaces on different line cards on LSRs.

A consideration in modifying the identity label (the MPLS label ordinarily used to identify the LSP, Virtual Private Network, Pseudowire etc) to indicate the batch is the impact that this has on the path chosen by the ECMP mechanism. When the member of the ECMP path set is chosen by deep packet inspection a change of batch represented by a change of identity label will have no impact on the ECMP path. Where the path member is chosen by reference to an entropy label [[RFC6790](#)] then provided that the entropy label is higher in the stack than the label that is changing the batch identifier again there will be no change to the chosen ECMP path. ECMP is so pervasive in multi-point to (multi-) point networks that some method of avoiding accounting errors introduced by ECMP needs to be supported.

#### [4.](#) Delay Measurement Considerations

Most of the existing delay measurement methods are active measurement that depend on the extra injected test packet to evaluate the delay of a path. With the active measurement method, the rate, numbers and interval between the injected packets may affect the accuracy of the results. Also, for injected test packets, these may not be co-routed with the data traffic due to ECMP. Thus there exists a requirements to measure the delay of the real traffic. For loss delay, the

identity considerations described in [Section 3](#) also apply.

## 5. Units of identification

The most basic unit of identification is the identity of the node processed the packet on its entry to the MPLS network. However, the required unit of identification may vary depending on the use case for accounting, performance measurement or other types of packet observations. In particular note that there may be a need to impose identify at several different layers of the MPLS label stack.

This document considers following units of identifications:

- o Per source LSR - everything from one source is aggregated.
- o Per group of LSPs chosen by an ingress LSR - an ingress LSP aggregates group of LSPs (ex: all LSPs of a tunnel).
- o Per LSP - the basic form.

- o Per flow [[RFC6790](#)] within an LSP - fine graining method.

Note that a finer grained identity resolution is needed when there is a need to perform these operations on a flow not readily identified by some other element in the label stack. Such fine grained resolution may be possible by deep packet inspection, but this may not always be possible, or it may be desired to minimise processing costs by doing only in entry to the network, and adding a suitable identifier to the packet for reference by other network elements. An example of such a fine grained case might be traffic from a specific application, or from a specific application from a specific source, particularly if matters related to service level agreement or application performance were being investigated.

We can thus characterize the identification requirement in the following broad terms:

- o There needs to be some way for an egress LSR to identify the ingress LSR with an appropriate degree of scope. This concept is discussed further in [Section 7](#).
- o There needs to be a way to identify a specific LSP at the egress

node. This allows for the case of instrumenting multiple LSPs operate between the same pair of nodes. In such cases the identity of the ingress LSR is insufficient.

- o In order to conserve resources such as labels, counters and/or compute cycles it may be desirable to identify an LSP group so that a operation can be performed on the group as an aggregate.
- o There needs to be a way to identify a flow within an LSP. This is necessary when investigating a specific flow that has been aggregated into an LSP.

The unit of identification and the method of determining which packets constitute a flow will be application or use-case specific and is out of scope of this memo.

## 6. Types of LSP

We need to consider a number of types of LSP. The two simplest types to monitor are point to point LSPs and point to multi-point LSPs. The ingress LSR for a point to point LSP, such as those created using the RSVP-TE signalling protocol, or those that conform to the MPLS-TP may be identified by inspection of the top label in the stack, since at any PE or P router on the path this is unique to the ingress-egress pair at every hop at a given layer in the LSP hierarchy. Provided that penultimate hop popping is disabled, the identity of

the ingress LSR of a point to point LSP is available at the egress LSR and thus determining the identity of the ingress LSR must be regarded as a solved problem. Note however that the identity of a flow cannot to be determined without further information.

In the case of a point to multi-point LSP the identity of the ingress LSR may also be inferred from the top label. [Editor's note - there was discussion of the following sentence amongst the authors and this needs to be looked at in the next version]. However, it may not possible to adequately from the top label alone. In designing any solution it is desirable that a common flow identity solution be used for both point to point and point to multi-point LSP types. Similarly it is desirable that a common method of LSP group identification be used.

[Editor's note: The following text was in -00, and a review comment asks why. At the time of editing I cannot remember the context. If the original authors cannot remember why by the next version, it will be deleted] In the above cases, an explicit non-null label is needed to provide context at the egress LSR. This is widely supported MPLS feature.

A more interesting case, and the core purpose of this memo, is the case of a multi-point to point LSP. In this case the same label is normally used by multiple ingress or upstream LSRs and hence source identification is not possible by inspection of the top label by egress LSRs. It is therefore necessary for a packet to be able to explicitly convey any of the identity types described in [Section 5](#).

Similarly, in the case of a multi-point to multi-point LSP the same label is normally used by multiple ingress or upstream LSRs and hence source identification is not possible by inspection of the top label by egress LSRs. The various types of identity described in [Section 5](#) are again needed. Note however, that the scope of the identity may be constrained to be unique within the set of multi-point to multi-point LSPs terminating on any common node.

## [7](#). Network Scope

The scope of identification can be constrained to the set of flows that are uniquely identifiable at an ingress LSR, or some aggregation thereof. There is no question of an ingress LSR seeking assistance from outside the MPLS protocol domain.

In any solution that constrains itself to carrying the required identity in the MPLS label stack rather than in some different associated data structure, constraints on the label stack size imply that the scope of identity reside within that MPLS domain. For

similar reasons the identity scope of a component of an LSP should be constrained to the scope of that LSP.

## [8](#). Backwards Compatibility

In any network it is unlikely that all LSRs will have the same capability to support the methods of identification discussed in this memo. It is therefore an important constraint on any identity

solution that it is backwards compatible with deployed MPLS equipment to the extent that deploying the new feature will not disable anything that currently works on a legacy equipment.

This is particularly the case when the deployment is incremental or when the feature is not required for all LSRs or all LSPs. Thus in broad the flow identification design MUST support the co-existence of LSRs that can and cannot identify the traffic components described in [Section 5](#). In addition the identification of the traffic components described in [Section 5](#) MUST be an optional feature that is disabled by default. As a design simplification, a solution MAY require that all egress LSRs of a point to multipoint or a multi-point to multipoint LSP support the identification type in use so that a single packet can be correctly processed by all egress devices. The corollary of this last point is that either all egress LSRs are enabled to support the required identity type, or none of them are.

## 9. Dataplane

There is a huge installed base of MPLS equipment, typically this type of equipment remains in service for an extended period of time, and in many cases hardware constraints mean that it is not possible to upgrade its dataplane functionality. Changes to the MPLS data plane are therefore expensive to implement, add complexity to the network, and may significantly impact the deployability of a solution that requires such changes. For these reasons, the MPLS designers have set a very high bar to changes to the MPLS data plane, and only a very small number have been adopted. Hence, it is important that the method of identification must minimize changes to the MPLS data plane. Ideally method(s) of identification that require no changes to the MPLS data plane should be given preferential consideration. If a method of identification makes a change to the data plane is chosen it will need to have a significant advantage over any method that makes no change, and the advantage of the approach will need to be carefully evaluated and documented. If a change is necessary to the MPLS data plane proves necessary, it should be (a) be as small a change as possible and (b) be a general purpose method so as to maximise its use for future applications. It is imperative that, as far as can be foreseen, any necessary change made to the MPLS data

plane does not impose any foreseeable future limitation on the MPLS



data plane.

Stack size is an issue with many MPLS implementations both as a result of hardware limitations, and due to the impact on networks and applications where a large number of small payloads need to be transported. In particular one MPLS payload may be carried inside another. For example one LSP may be carried over another LSP, or a PW or similar multiplexing construct may be carried over an LSP and identification may be required at both layers. Of particular concern is the implementation of low cost edge LSRs that for cost reasons have a significant limit on the number of Label Stack Elements (LSEs) that they can impose or dispose. Therefore, any method of identity MUST NOT consume an excessive number of unique labels, and MUST NOT result in an excessive increase in the size of the label stack.

The MPLS data plane design provides two types of special purpose labels: the original 16 reserved labels and the much larger set of special purpose labels defined in [\[RFC7274\]](#). The original reserved labels need one LSE, and the newer [\[RFC7274\]](#) special purpose labels need two LSEs. Given the tiny number of original reserved labels, it is core to the MPLS design philosophy that this scarce resource is only used when it is absolutely necessary. Using a single LSE reserved or special purpose label to encode flow identity thus requires two stack entries, one for the reserved label and one for the flow identity. The larger set of [\[RFC7274\]](#) labels requires two labels stack entries for the special purpose label itself and hence a total of three label stack entries to encode the flow identity.

The use of special purpose labels (SPL) [\[RFC7274\]](#) as part of a method to encode the identity information therefore has a number of undesirable implications for the data plane and hence whilst a solution may use SPL(s), methods that do not require SPLs need to be carefully considered.

## [10.](#) Control Plane

Any flow identity design should both seek to minimise the complexity of the control plane and should minimise the amount of label coordination needed amongst LSRs.

## [11.](#) Manageability Considerations

This will be provided in a future version of this document.

## 12. Privacy Considerations

The inclusion of originating and/or flow information in a packet provides more identity information and hence potentially degrades the privacy of the communication. Recent IETF concerns on pervasive monitoring would lead it to prefer a solution that does not degrade the privacy of user traffic below that of an MPLS network not implementing the flow identification feature. The minimizing the scope of the identity indication can be useful in minimizing the observability of the flow characteristics.

## 13. Security Considerations

Any solution to the flow identification needs must not degrade the security of the MPLS network below that of an equivalent network not deploying the specified identity solution. Propagation of identification information outside the MPLS network imposing it must be disabled by default. Any solution should provide for the restriction of the identity information to those components of the network that need to know it. It is thus desirable to limit the knowledge of the identify of an endpoint to only those LSRs that need to participate in traffic flow.

## 14. IANA Considerations

EDITOR'S NOTE: This section may be removed on publication

This memo has no IANA considerations.

## 15. Acknowledgements

The authors thank Nobo Akiya (nobo@cisco.com), Nagendra Kumar Nainar (naikumar@cisco.com) and George Swallow (swallow@cisco.com) for their comments.

## 16. References

### 16.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.

### 16.2. Informative References

---

Internet-Draft

MPLS FI

March 2015

[I-D.tempia-ippm-p3m]

Capello, A., Cociglio, M., Fioccola, G., Castaldelli, L., and A. Bonda, "A packet based method for passive performance monitoring", [draft-tempia-ippm-p3m-00](#) (work in progress), March 2015.

[RFC6374] Frost, D. and S. Bryant, "Packet Loss and Delay Measurement for MPLS Networks", [RFC 6374](#), September 2011.

[RFC6790] Kompella, K., Drake, J., Amante, S., Henderickx, W., and L. Yong, "The Use of Entropy Labels in MPLS Forwarding", [RFC 6790](#), November 2012.

[RFC7274] Kompella, K., Andersson, L., and A. Farrel, "Allocating and Retiring Special-Purpose MPLS Labels", [RFC 7274](#), June 2014.

#### Authors' Addresses

Stewart Bryant  
Cisco Systems

Email: [stbryant@cisco.com](mailto:stbryant@cisco.com)

Carlos Pignataro  
Cisco Systems

Email: [cpignata@cisco.com](mailto:cpignata@cisco.com)

Mach Chen  
Huawei

Email: [mach.chen@huawei.com](mailto:mach.chen@huawei.com)

Zhenbin Li  
Huawei

Email: lizhenbin@huawei.com

Gregory Mirsky  
Ericsson

Email: gregory.mirsky@ericsson.com

Bryant, et al.

Expires September 27, 2015

[Page 10]