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

S. Bryant
C. Pignataro
Cisco Systems

October 24, 2014

MPLS Flow Identification
draft-bryant-mpls-flow-ident-00

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 April 27, 2015.

Copyright Notice

Table of Contents

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]. However this work 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 is difficult to assess a loss using 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. 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.

2. Loss Measurement Considerations

Modern networks normally drop very few packets, thus packet loss measurement are highly sensitive to counter errors. Without some form of coloring or batch marking such as that proposed in [I-D.tempia-opsawg-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 advantage 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. The packet accounting mechanism is then able to operate on a 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 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 colour 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 colour 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.

3. 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 mat be a need to impose identify at several different layers of the MPLS label stack.

This document considers follow unit of identifications:

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

- 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 5.
- 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.
- 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.
- 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 method of determining which packets constitute a flow is outside the scope of this memo.

4. 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. However it is not possible to identify a flow from the top label, nor is it possible to directly identify the ingress LSR since there may be many point to multi-point LSP originating at that LSR. 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.

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 3.

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 3 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.

Any method of identity must not consume an excessive number of unique labels, nor result in an excessive increase in the size of the label stack (Section 7).

5. 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 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.

6. 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 3). In addition the identification of the traffic components described in Section 3 needs to 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 to 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.

7. 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 LSEs that they can impose or dispose.

The MPLS data plane design provides only a tiny number of reserved labels, it is therefore 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. A larger special purpose labels space is available [RFC7274] but this requires two labels stack entries for the reserved 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.

8. Control Plane

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

9. Manageability Considerations

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

10. 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.

11. 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.

12. IANA Considerations

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

This memo has no IANA considerations.

13. Acknowledgements

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

14. References

14.1. Normative References

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

14.2. Informative References

[I-D.tempia-opsawg-p3m]
Capello, A., Cociglio, M., Castaldelli, L., and A. Bonda, "A packet based method for passive performance monitoring", draft-tempia-opsawg-p3m-04 (work in progress), February 2014.

[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

Carlos Pignataro
Cisco Systems

Email: cpignata@cisco.com