MPLS Working Group   S. Bryant
Internet-Draft   Huawei
Intended status: Informational   C. Pignataro
Expires: September 2, 2018   Cisco Systems
                                                          M. Chen
                                                            Z. Li
                                                           Huawei
                                                        G. Mirsky
                                                        ZTE Corp.
                                                   March 01, 2018

MPLS Flow Identification Considerations
draft-ietf-mpls-flow-ident-07

Abstract

   This document discusses the aspects that need to be be considered
   when developing a solution for MPLS flow identification. The key
   application that needs this is in-band performance monitoring of MPLS
   flows when MPLS is used to encapsulate 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 https://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 2, 2018.

Copyright Notice

Table of Contents

1. Introduction

   This document discusses the aspects that need to be considered when
   developing a solution for MPLS flow identification. The key
   application that needs this is in-band performance monitoring of MPLS
   flows when MPLS is used for the encapsulation of user data packets.

   There is a need to identify flows in MPLS networks for various
   applications such as determining 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 injected Operations, Administration, and
   Maintenance (OAM) packets 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, [RFC6374] did not support different granularities of
   flow or address a number of multi-point cases in which two or more
   ingress Label Switching Routers (LSRs) could send packets to one or
   more destinations.

Improvements in link and transmission technologies have made it more
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, means that network operators now 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 measurement process
leaks 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. Loss Measurement Considerations

Modern networks, if not oversubscribed, generally drop relatively few
packets, thus packet loss measurement is highly sensitive to the
common demarcation of the exact set of packets to be measured for
loss. Without some form of coloring or batch marking such as that
proposed in [RFC8321] it may not be possible to achieve the required
accuracy in the loss measurement of customer data traffic. Thus
where accurate measurement of packet loss is required, it may be
economically advantageous, or even a technical requirement, to
include some form of marking in the packets to assign each packet to
a particular counter for loss measurement purposes.

Where this level of accuracy is required and the traffic between a
source-destination pair is subject to Equal-Cost Multipath (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 Label
Switched Paths (LSPs) subjected to ECMP because the network transit
time will be different for the various ECMP paths since:

1. The packets may traverse different sets of LSRs.

2. The packets may depart from different interfaces on different
    line cards on LSRs.

   3. The packets may arrive at different interfaces on different line
       cards on LSRs.

   A consideration with this solution on 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 changing the batch identifier will not
   result in a 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. Delay Measurement Considerations

   Most of the existing delay measurement methods are active methods
   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. Due to ECMP (or link aggregation techniques) injected test
   packets may traverse different links from the ones used by the data
   traffic. Thus there exists a requirement to measure the delay of the
   real traffic.

   For combined loss-delay measurements, both the loss and the delay
   considerations apply.

4. Units of identification

   The most basic unit of identification is the identity of the node
   that 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
   identity 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 grained method.

Note that a fine 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. However, this may not always be possible, or it may be
desired to minimize processing costs by doing this only on entry to
the network. Adding a suitable identifier to the packet for
reference by other network elements minumises the processing needed
by other network elements. An example of such a fine grained case
might be traffic belonging to a certain service or 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 6.

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
   operating 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 document.

5. 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 Resource Reservation Protocol - Traffic Engineering (RSVP-TE)
[RFC5420] Signaling protocol, or those that conform to the MPLS
Transport Profile (MPLS-TP) [RFC5654] may be identified by inspection
of the top label in the stack, since at any provider-edge (PE) or
provider (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 being carried in the
packet, or gleaned from some aspect of the packet payload.

In the case of a point to multi-point LSP, and in the absence of
Penultimate Hop Popping (PHP) the identity of the ingress LSR may
also be inferred from the top label. However, it may not possible to
adequately identify the flow from the top label alone, and thus
further information may need to be carried in the packet, or gleaned
from some aspect of the packet payload. 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, a context label [RFC5331] needs to be used to
provide the required identity information. This is widely supported
MPLS feature.

A more interesting case 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 the egress LSRs. It is therefore
necessary for a packet to be able to explicitly convey any of the
identity types described in Section 4.

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 4
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.

6. 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 need 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 choice of label and
label stack size imply that the scope of identity resides within that
MPLS domain. For similar reasons the identity scope of a component
of an LSP is constrained to the scope of that LSP.

7. 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
   document. It is therefore an important constraint on any flow
   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
   the feature is not required for all LSRs or all LSPs. Thus, the flow
   identification design must support the co-existence of both LSRs that
   can, and cannot, identify the traffic components described in
   Section 4. In addition the identification of the traffic components
   described in Section 4 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 multi-point or a multi- point to multi-
   point 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.

8. 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, MPLS users 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 maximize 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.

9. 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.

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 [RFC7258] 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 choice of using
MPLS technology for this OAM solution has a privacy advantage as the
choice of the label identifying a flow is limited to the scope of the
MPLS domain and does not have any dependency on the user data's

identification. This minimizes 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.
In order to preserve present assumptions about MPLS privacy
properties, 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. The choice
of using MPLS technology for this OAM solution, with MPLS
encapsulation of user traffic, provides for a key advantage over
other data plane solutions, as it provides for a controlled access
and trusted domain within a Service Provider's network.

For a more comprehensive discussion of MPLS security and attack
mitigation techniques, please see the Security Framework for MPLS and
GMPLS Networks [RFC5920].

12. IANA Considerations

This document has no IANA considerations. (At the discretion of the
RFC Editor this section may be removed before publication).

13. Acknowledgments

The authors thank Nobo Akiya, Nagendra Kumar Nainar, George Swallow
and Deborah Brungard for their comments.

14. Informative References

[RFC5331] Aggarwal, R., Rekhter, Y., and E. Rosen, "MPLS Upstream
           Label Assignment and Context-Specific Label Space",
           RFC 5331, DOI 10.17487/RFC5331, August 2008,
           <https://www.rfc-editor.org/info/rfc5331>.

[RFC5420] Farrel, A., Ed., Papadimitriou, D., Vasseur, JP., and A.
           Ayyangarps, "Encoding of Attributes for MPLS LSP
           Establishment Using Resource Reservation Protocol Traffic
           Engineering (RSVP-TE)", RFC 5420, DOI 10.17487/RFC5420,
           February 2009, <https://www.rfc-editor.org/info/rfc5420>.

   [RFC5654] Niven-Jenkins, B., Ed., Brungard, D., Ed., Betts, M., Ed.,
              Sprecher, N., and S. Ueno, "Requirements of an MPLS
              Transport Profile", RFC 5654, DOI 10.17487/RFC5654,
              September 2009, <https://www.rfc-editor.org/info/rfc5654>.

   [RFC5920] Fang, L., Ed., "Security Framework for MPLS and GMPLS
              Networks", RFC 5920, DOI 10.17487/RFC5920, July 2010,
              <https://www.rfc-editor.org/info/rfc5920>.

   [RFC6374] Frost, D. and S. Bryant, "Packet Loss and Delay
              Measurement for MPLS Networks", RFC 6374,
              DOI 10.17487/RFC6374, September 2011,
              <https://www.rfc-editor.org/info/rfc6374>.

   [RFC6790] Kompella, K., Drake, J., Amante, S., Henderickx, W., and
              L. Yong, "The Use of Entropy Labels in MPLS Forwarding",
              RFC 6790, DOI 10.17487/RFC6790, November 2012,
              <https://www.rfc-editor.org/info/rfc6790>.

   [RFC7258] Farrell, S. and H. Tschofenig, "Pervasive Monitoring Is an
              Attack", BCP 188, RFC 7258, DOI 10.17487/RFC7258, May
              2014, <https://www.rfc-editor.org/info/rfc7258>.

   [RFC7274] Kompella, K., Andersson, L., and A. Farrel, "Allocating
              and Retiring Special-Purpose MPLS Labels", RFC 7274,
              DOI 10.17487/RFC7274, June 2014,
              <https://www.rfc-editor.org/info/rfc7274>.

   [RFC8321] Fioccola, G., Ed., Capello, A., Cociglio, M., Castaldelli,
              L., Chen, M., Zheng, L., Mirsky, G., and T. Mizrahi,
              "Alternate-Marking Method for Passive and Hybrid
              Performance Monitoring", RFC 8321, DOI 10.17487/RFC8321,
              January 2018, <https://www.rfc-editor.org/info/rfc8321>.

Authors' Addresses

   Stewart Bryant
   Huawei

   Email: stewart.bryant@gmail.com

   Carlos Pignataro
   Cisco Systems

   Email: cpignata@cisco.com

   Mach Chen
   Huawei

   Email: mach.chen@huawei.com

   Zhenbin Li
   Huawei

   Email: lizhenbin@huawei.com

   Gregory Mirsky
   ZTE Corp.

   Email: gregimirsky@gmail.com