Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: January 16, 2014

M. Chen
X. Xu
Z. Li
Huawei
L. Fang
Cisco

July 15, 2013

MultiProtocol Label Switching (MPLS) Source Label
draft-chen-mpls-source-label-00

Abstract

An MultiProtocol Label Switching (MPLS) label is originally defined to identify a Forwarding Equivalence Class (FEC), a packet is assigned to a specific FEC based on its network layer destination address. It's difficult or even impossible to derive the source information from the label. For some applications, source identification is a critical requirement. For example, performance monitoring, traffic matrix measurement and collection, where the monitoring node needs to identify where a packet was sent from.

This document introduces the concept of Source Label (SL) that is carried in the label stack and used to identify the ingress Label Switching Router (LSR) of an Label Switched Path (LSP).

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 RFC 2119 [RFC2119].

This Internet-Draft will expire on January 16, 2014.

Copyright Notice

Table of Contents

1. Problem Statement and Introduction

An MultiProtocol Label Switching (MPLS) label [RFC3031] is originally defined for packet forwarding and assumes the forwarding/destination address semantics. As no source address information is carried in the label stack, there is no way to directly derive the source address information from the label or label stack.

MPLS LSPs can be categorized into four different types:

Point-to-Point (P2P)

Point-to-Multipoint (P2MP)

Multipoint-to-Point (MP2P)

Multipoint-to-Multipoint (MP2MP)

LSPs that are established by the Resource Reservation Protocol Traffic Engineering (RSVP-TE) [RFC3209] and Pseudowires (PWs) belong to P2P or P2MP types. LSPs that are established by the classic Label Distribution Protocol (LDP) [RFC5036], Layer 3 Private Network (L3VPN) and Virtual Local Area Network (VPLS) LSPs belong to MP2P or MP2MP types.

For those LSPs belong to the MP2P and MP2MP types, it is not possible to derive the source address information from the label. For the P2P or P2MP LSPs, the source address information may be implicitly derived from the label (e.g., P2P or P2MP LSPs established by RSVP-TE) , but it requires that some further information is used (e.g., control plane information). However, this is not always possible for all P2P LSPs. One example is the Multi-Segment Pseudowire (MS-PW), it is impossible to derive the source address information from the PW label. Because an MS-PW label assumes the forwarding and destination address semantics which is quite different from the source address semantics that a Single-Segment Pseudowire (SS-PW) label assumes.

Comparing to the pure IP forwarding where both source and destination addresses are encoded in the IP packet header, the essential issue of the MPLS encoding is that the label stack does not explicitly include any source address information, i.e., a Source Label (SL). For some applications, source identification is a critical requirement. For example, performance monitoring, the monitoring nodes need to identify where packets were sent from and then can count the packets according to some constraints. In addition, traffic matrix measurement and collection is the precondition of traffic steering, and capable of traffic steering is an important requirement of Software Defined Network (SDN). To measure and collect traffic matrix information, the source address information is necessary.

This document introduces the concept of a Source Label. A SL uniquely identifies a node within an administrative domain, it is carried in the label stack and used to identify the ingress LSR(s) of an LSP.

2. Source Label

A Source Label is defined to uniquely identify a node that is (one of) the ingress LSR(s) to a specific LSP. In its function as a Source Label (ingress node identifier), it MUST be unique within a domain. In cases where a Source Label is used across domains it MUST be unique within the scope it is used. How to guarantee the uniqueness of Source Labels is out of scope for this document. Source Labels are not used for forwarding.

In order to indicate whether a label is a source label, a Source Label Indicator (SLI) is introduced. The SLI is a reserved label that is placed immediately before the source label in the label stack, which is used to indicate that the next label in the label stack is a source label. The value of SLI is TBD1.

3. Use Cases

3.1. Performance Measurement

There are two typical types of performance measurement: one is active performance measurement, and the other is passive performance measurement.

In active performance measurement the receiver measures the injected packets to evaluate the performance of a path. The active measurement measures the performance of the extra injected packets. The IP Performance Metrics (IPPM) working group has defined specifications for the active performance measurement.

In passive performance measurement, no artificial traffic is injected into the flow and measurements are taken to record the performance metrics of the real traffic. The Multiprotocol Label Switching (MPLS) PM protocol [RFC6374] for packet loss is an example of passive performance measurement. For a specific receiver, in order to count the received packets of a flow, it has to know whether a received packet belongs to which target flow under test and the source identification is a critical condition.

As discussed in the previous section, the existing MPLS label or label stack do not carry the source information. So, for an LSP, the ingress LSR can put a source label in the label stack, and then the egress LSR can use the source label for packets identifying and counting.

3.2. Traffic Matrix Measurement and Steering

A Traffic Matrix (TM) provides, for every ingress node (i) into the network and every egress node (j) out of the network, the volume of traffic T(i,j) from i to j over a given time interval.

Since the ingress node knows the source and destination of the traffic, it's normal to measure the traffic matrix at every ingress node. But in some scenarios, it may need to measure the traffic at the egress or intermediate nodes. Taking Figure 1 as an example, from the west to east point of view, there are three ingress nodes (I1, I2 and I3) and three egress nodes (E1, E2 and E3), A, B and C are intermediate nodes. It is not necessary to measure the traffic matrix of the whole network all the time, it sometimes just wants to know the received traffic matrix of a specific egress node (e.g., E2). So, to measure received traffic matrix at node E2 would be then a better choice.

In addition, for an intermediate node (e.g., A), it may need to measure the transmitted traffic hence to steer some traffic from the congestion path to idle path.

Wherever at egress or intermediate node, source identification is necessary. The ingress LSR can put the source label into the label stack to help the egress and intermediate LSR to identify and measure the traffic.

```
           +--+
           |I1|
           +--+            +--+    +----+
              \           /|B |----|E1 |
               \         / +--+    +----+
                \       /      \
```

```
                 \     /        \
                  \   /          \+--+
       +--+        \+--+         /|E2 |
       |I2|--------/|A |        / +--+
       +--+       / +--+       /
                 /     \      /
                /       \+--+/    +--+
               /         |C |-----|E3|
          +--+/          +--+     +--+
          |I3|
          +--+
```

Figure 1: Traffic Matrix Measurement and Steering

4. Data Plane Processing

4.1. Ingress LSR

For an LSP, the ingress LSR MUST make sure that the egress LSR is able to process the Source Label before inserting a SL and SLI into the label stack. Therefore, an egress LSR SHOULD signal (see Section 5.1) to the ingress LSR whether it is able to process the Source Label. Once the ingress LSR knows that the egress LSR can process Source Label, it can choose whether or not to insert the SL and SLI into the label stack.

When a SL to be included in a label stack, the steps are as follows:

1. Push the SL label, the Bos bit for the SL depends on whether the SL is the bottom label;

2. Push the SLI, the TTL and TC field for the SLI SHOULD be set to the same values as for the LSP Label (L);

3. Push the LSP Label (L) .

Then the label stack looks like: <...L, SLI, SL...>. There may be multiple pairs of SLI and SL inserted into the label stack, each pair is related to an LSP. For an LSP, only one pair of SLI and SL SHOULD be inserted.

4.2. Transit LSR

There is no change in forwarding behavior for transit LSRs. But if a transit LSR can recognize the SLI, it may use the SL to collect traffic throughput and/or measure the performance of the LSP.

4.3. Egress LSR

When an egress LSR receives a packet with a SLI/SL pair, if the egress LSR is able to process the SL; it pops the LSP label (if have), SLI and SL; then processes remaining packet header as normal. If the egress LSR is not able to process the SL, the packet SHOULD be dropped.

4.4. Penultimate Hop LSR

The penultimate hop LSR MUST not pop the SLI and SL.

5. Source Label Signaling

Source label signaling includes two aspects: one is source label capability signaling, the other is source label distribution.

5.1. Source Label Capability Signaling

Before inserting a source label in the label stack, an ingress LSR MUST know whether the egress LSR is able to process the source label. Therefore, an egress LSR should signal to the ingress LSRs its ability to process the Source Label. This is called Source Label Capability (SLC), it is very similar to the "Entropy Label Capability (ELC)"[RFC6790].

5.1.1. LDP Extensions

A new LDP TLV [RFC5036], SLC TLV, is defined to signal an egress's ability to process source label. The SLC TLV may appear as an Optional Parameter of the Label Mapping Message. The presence of the SLC TLV in a Label Mapping Message indicates to ingress LSRs that the egress LSR can process source labels for the associated LSP.

The structure of the SLC TLV is shown below.

```
  0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|U|F|        Type (TBD2)        |            Length (0)         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 1: Source Label Capability TLV

This U bit MUST be set to 1. If the SLC TLV is not understood by the receiver, then it MUST be ignored.

This F bit MUST be set to 1. Since the SLC TLV is going to be propagated hop-by-hop, it should be forwarded even by nodes that may not understand it.

Type: TBD2.

Length field: This field specifies the total length in octets of the SLC TLV and is defined to be 0.

An LSR that receives a Label Mapping with the SLC TLV but does not understand it MUST propagate it intact to its neighbors and MUST NOT send a notification to the sender (following the meaning of the U- and F-bits). An LSR X may receive multiple Label Mappings for a given FEC F from its neighbors. In its turn, X may advertise a Label Mapping for F to its neighbors. If X understands the SLC TLV, and if any of the advertisements it received for FEC F does not include the SLC TLV, X MUST NOT include the SLC TLV in its own advertisements of F. If all the advertised Mappings for F include the SLC TLV, then X MUST advertise its Mapping for F with the SLC TLV. If any of X's neighbors resends its Mapping, sends a new Mapping or sends a Label Withdraw for a previously advertised Mapping for F, X MUST re-evaluate the status of SLC for FEC F, and, if there is a change, X MUST re-advertise its Mapping for F with the updated status of SLC.

5.1.2. BGP Extensions

When Border Gateway Protocol (BGP) [RFC4271] is used for distributing Network Layer Reachability Information (NLRI) as described in, for example, [RFC3107], [RFC4364], the BGP UPDATE message may include the SLC attribute as part of the Path Attributes. This is an optional, transitive BGP attribute of value TBD3. The inclusion of this attribute with an NLRI indicates that the advertising BGP router can process source labels as an egress LSR for all routes in that NLRI.

A BGP speaker S that originates an UPDATE should include the SLC attribute only if both of the following are true:

A1: S sets the BGP NEXT_HOP attribute to itself AND

A2: S can process source labels.

Suppose a BGP speaker T receives an UPDATE U with the SLC attribute. T has two choices. T can simply re-advertise U with the SLC attribute if either of the following is true:

B1: T does not change the NEXT_HOP attribute OR

B2: T simply swaps labels without popping the entire label stack and processing the payload below.

An example of the use of B1 is Route Reflectors. However, if T changes the NEXT_HOP attribute for U and in the data plane pops the entire label stack to process the payload, T MAY include an SLC attribute for UPDATE U' if both of the following are true:

C1: T sets the NEXT_HOP attribute of U' to itself AND

C2: T can process source labels. Otherwise, T MUST remove the SLC attribute.

5.1.3. RSVP-TE Extensions

Source label support is signaled in RSVP-TE [RFC3209] using the Source Label Capability (SLC) flag in the Attribute Flags TLV of the LSP_ATTRIBUTES object [RFC5420]. The presence of the SLC flag in a Path message indicates that the ingress can process entropy labels in the upstream direction; this only makes sense for a bidirectional LSP and MUST be ignored otherwise. The presence of the SLC flag in a Resv message indicates that the egress can process entropy labels in the downstream direction. The bit number for the SLC flag is TBD4.

5.2. Source Label Distribution

Based on the Source Label, an egress or intermediate LSR can identify from where an MPLS packet is sent. To achieve this, the egress and/or intermediate LSRs have to know which ingress LSR is related to which Source Label before using the Source Label to derive the source information. Therefore, there needs a mechanism to distribute the mapping information between an ingress LSR and its Source Label. For example, defines extensions to LDP, BGP, RSVP-TE and/or Interior Gateway Protocol (IGP) to distribute to source label. The source label distribution will be defined in future revision or another document.

6. IANA Considerations

6.1. Source Label Indication

IANA is required to allocate a reserved label (TBD1) for the Source Label Indicator (SLI) from the "Multiprotocol Label Switching Architecture (MPLS) Label Values" Registry.

6.2. LDP Source Label Capability TLV

IANA is required to allocate a value of TBD2 from the IETF Consensus range (0x0001-0x07FF) in the "TLV Type Name Space" registry as the "Source Label Capability TLV".

6.3. BGP Source Label Capability Attribute

IANA is required to allocate a Path Attribute Type Code TBD3 from the "BGP Path Attributes" registry as the "BGP Source Label Capability Attribute".

6.4. RSVP-TE Source Label Capability

IANA is required to allocate a new bit from the "Attribute Flags" sub-registry of the "Resource Reservation Protocol-Traffic Engineering (RSVP-TE) Parameters" registry.

Bit No	Name	Attribute Flags Path	Attribute Flags Resv	RRO
TBD4	Source Label Capability	Yes	Yes	No

7. Security Considerations

TBD.

8. Acknowledgements

The process of "Source Label Capability Signaling" is largely referred to the process of "ELC signaling"[RFC6790].

The authors would like to thank Carlos Pignataro, Loa Andersson for their review, suggestion and comments to this document.

9. References

9.1. Normative References

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

[RFC3031] Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol Label Switching Architecture", RFC 3031, January 2001.

[RFC3107] Rekhter, Y. and E. Rosen, "Carrying Label Information in BGP-4", RFC 3107, May 2001.

[RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, December 2001.

[RFC4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 4364, February 2006.

[RFC5036] Andersson, L., Minei, I., and B. Thomas, "LDP Specification", RFC 5036, October 2007.

[RFC5420] Farrel, A., Papadimitriou, D., Vasseur, JP., and A. Ayyangarps, "Encoding of Attributes for MPLS LSP Establishment Using Resource Reservation Protocol Traffic Engineering (RSVP-TE)", RFC 5420, February 2009.

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

9.2. Informative References

[RFC4271] Rekhter, Y., Li, T., and S. Hares, "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, January 2006.

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

Authors' Addresses

Mach(Guoyi) Chen
Huawei

Email: mach.chen@huawei.com

Xiaohu Xu
Huawei

Email: xuxiaohu@huawei.com

Zhenbin Li
Huawei

Email: lizhenbin@huawei.com

Luyuan Fang
Cisco

Email: lufang@cisco.com