MPLS Working Group
Internet-Draft
Intended status: Standards Track
Expires: May 29, 2021

R. Gandhi, Ed.
Z. Ali
C. Filsfils
F. Brockners
Cisco Systems, Inc.
B. Wen
V. Kozak
Comcast

November 25, 2020

MPLS Data Plane Encapsulation for In-situ OAM Data
draft-gandhi-mpls-ioam-sr-04

Abstract

In-situ Operations, Administration, and Maintenance (IOAM) records operational and telemetry information in the data packet while the packet traverses a path between two nodes in the network. This document defines how IOAM data fields are transported using the MPLS data plane encapsulation, including Segment Routing (SR) with MPLS data plane (SR-MPLS).

Table of Contents

1. Introduction

In-situ Operations, Administration, and Maintenance (IOAM) records operational and telemetry information within the packet while the packet traverses a particular network domain. The term "in-situ" refers to the fact that the IOAM data fields are added to the data packets rather than being sent within the probe packets specifically dedicated to OAM or Performance Measurement (PM). The IOAM data fields are defined in [I-D.ietf-ippm-ioam-data], and can be used for various use-cases for OAM and PM. The IOAM data fields are further updated in [I-D.ietf-ippm-ioam-direct-export] for direct export use-cases and in [I-D.ietf-ippm-ioam-flags] for Loopback and Active flags.

This document defines how IOAM data fields are transported using the MPLS data plane encapsulations, including Segment Routing (SR) with MPLS data plane (SR-MPLS).

2. Conventions

2.1. Requirement 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] [RFC8174] when, and only when, they appear in all capitals, as shown here.

2.2. Abbreviations

Abbreviations used in this document:

ECMP Equal Cost Multi-Path

IOAM In-situ Operations, Administration, and Maintenance

MPLS Multiprotocol Label Switching

OAM Operations, Administration, and Maintenance

PM Performance Measurement

POT Proof-of-Transit

PSID Path Segment Identifier

SR Segment Routing

SR-MPLS Segment Routing with MPLS Data plane

3. IOAM Data Field Encapsulation in MPLS Header

The IOAM data fields defined in [I-D.ietf-ippm-ioam-data] are used. IOAM data fields are carried in the MPLS header as shown in Figure 1. More than one trace options can be present in the IOAM data fields. The IOAM Indicator Label is added at the bottom of the MPLS label stack (S flag set to 1) and it indicates the presence of the IOAM data field(s) in the MPLS header.

The data packets with IOAM data fields carry only one IOAM Indicator Label in the MPLS header. Any intermediate node that adds additional MPLS encapsulation in the MPLS header may further update the IOAM data fields in the header without inserting another IOAM Indicator Label.

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Extension Label (15)                 | TC  |0|  TTL          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  IOAM Indicator Label                 | TC  |1|  TTL          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+<-+
|0 0 1 0|R|R|R|R| Block Number  | IOAM-OPT-Type |IOAM HDR Length|  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+  I
|                                                               |  O
|                                                               |  A
~                 IOAM Option and Data Space                    ~  M
|                                                               |  |
|                                                               |  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+<-+
|                                                               |
|                                                               |
|                 Payload + Padding                             |
|                                                               |
|                                                               |
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 1: IOAM Encapsulation in MPLS Header

IOAM Indicator Label (IIL) (Edge-to-Edge or Hob-By-Hop) as defined in this document.

The fields related to the encapsulation of IOAM data fields in the MPLS header are defined as follows:

IP Version Number 0010b: The IP Version Number Field 0010b allows to avoid incorrect IP header-based hashing over ECMP paths that uses the value 0x4 (for IPv4) and value 0x6 (for IPv6) [RFC4928].

Block Number: The Block Number can be used to aggregate the IOAM data collected in data plane, e.g. compute measurement metrics for each block of a flow. It is also used to correlate the IOAM data on different nodes.

R Bits: Reserved Bits MUST be set to zero upon transmission and ignored upon receipt.

IOAM-OPT-Type: 8-bit field defining the IOAM Option type, as defined in Section 8.1 of [I-D.ietf-ippm-ioam-data].

IOAM HDR LEN: 8-bit unsigned integer. Length of the IOAM HDR in 4-octet units.

IOAM Option and Data Space: IOAM option header and data is present as defined by the IOAM-OPT-Type field, and is defined in Section 5 of [I-D.ietf-ippm-ioam-data].

3.1. IOAM Indicator Labels

IOAM Indicator Label is used to indicate the presence of the IOAM data field in the MPLS header.

Different IOAM Indicator Labels are used for E2E and HbH IOAM to optimize processing on transit nodes and for checking if IOAM data fields need to be processed. If only edge nodes need to process IOAM data then E2E IOAM Indicator Label is used so that transit nodes can ignore it. If both edge and transit nodes need to process IOAM data then HbH IOAM Indicator Label is used.

The SR path computation needs to know the Maximum SID Depth (MSD) that can be imposed at each node/link of a given SR path [RFC8664]. This ensures that the SID stack depth of a computed path does not exceed the number of SIDs the node is capable of imposing. The MSD used for path computation MUST include the IOAM Indicator Label.

4. Procedure for Edge-to-Edge IOAM

The Edge-to-Edge (E2E) IOAM includes IOAM Option-Type as Edge-to-Edge Option-Type [I-D.ietf-ippm-ioam-data]. This section summarizes the procedure for data encapsulation and decapsulation for Edge-to-Edge IOAM in MPLS header.

- o The encapsulating node inserts the E2E IOAM Indicator Label and one or more IOAM data field(s) in the MPLS header.

- o The decapsulating node "forwards and punts the timestamped copy" of the data packet including IOAM data fields when the node recognizes the E2E IOAM Indicator Label. The copy of the data packet is punted to the slow path for OAM processing and is not necessarily punted to the control-plane. The receive timestamp is required by various E2E OAM use-cases.

- o The decapsulating node processes the IOAM data field(s) using the procedures defined in [I-D.ietf-ippm-ioam-data]. An example of IOAM processing may be to export the data fields, send data fields via Telemetry, etc.

- o The decapsulating node also pops the E2E IOAM Indicator Label and the IOAM data fields from the MPLS header.

4.1. Edge-to-Edge IOAM Indicator Label Allocation

E2E IOAM Indicator Label is used to indicate the presence of the E2E IOAM data field in the MPLS header. The E2E IOAM Indicator Label can be allocated using one of the following methods:

- o Labels assigned by IANA with value TBA1 and TBA2 from the Extended Special-Purpose MPLS Values [I-D.ietf-mpls-spl-terminology].

- o Labels allocated by a Controller from the global table of the decapsulating node. The Controller provisions the label on both encapsulating and decapsulating nodes.

- o Labels allocated by the decapsulating node and signalled or advertised in the network. The signaling and/or advertisement extension for this is outside the scope of this document.

5. Procedure for Hop-by-Hop IOAM

The Hop-by-Hop (HbH) IOAM includes IOAM Option-Types IOAM Pre-allocated Trace Option-Type, IOAM Incremental Trace Option-Type and IOAM Proof of Transit (POT) Option-Type [I-D.ietf-ippm-ioam-data]. This section summarizes the procedure for data encapsulation and decapsulation for Hop-by-hop IOAM in MPLS header.

- o The encapsulating node inserts the HbH IOAM Indicator Label and one or more IOAM data field(s) in the MPLS header.

- o The intermediate and decapsulating node enabled with IOAM functions "forwards and punts the timestamped copy" of the data packet including IOAM data fields when the node recognizes the HbH IOAM Indicator Label. The copy of the data packet is punted to the slow path for OAM processing and is not necessarily punted to the control-plane. The receive timestamp is required by various hop-by-hop OAM use-cases.

- o The intermediate and decapsulating node processes the IOAM data field(s) using the procedures defined in [I-D.ietf-ippm-ioam-data]. An example of IOAM processing may be to export the data fields, send data fields via Telemetry, etc.

- o The decapsulating node pops the HbH IOAM Indicator Label and the IOAM data fields from the MPLS header.

5.1. Hop-by-Hop IOAM Indicator Label Allocation

HbH IOAM Indicator Label is used to indicate the presence of the HbH IOAM data field in the MPLS header. The HbH IOAM Indicator Label can be allocated using one of the following methods:

- o Labels assigned by IANA with value TBA2 from the Extended Special-Purpose MPLS Values [I-D.ietf-mpls-spl-terminology].

- o Labels allocated by a Controller from the network-wide global table. The Controller provisions the labels on all nodes participating in IOAM functions along the data traffic path.

6. Considerations for ECMP

The encapsulating node needs to make sure the IOAM data field does not start with a well known IP Version Number (e.g. 0x4 for IPv4 and 0x6 for IPv6) as it can alter the hashing function for ECMP that uses the IP header. This is achieved by using the IOAM Indicator Label that followed by IP Version Number 0010b. This approach is consistent with utilizing 0000b or 0001b as the first nibble for IP Version Number after the MPLS label stack, as described in [RFC4928] [RFC4385].

Note that the hashing function for ECMP that uses the labels from the MPLS header may now include the IOAM Indicator Label.

When entropy label [RFC6790] is used for hashing function for ECMP, the procedure defined in this document does not alter the hashing function.

7. Node Capability

The decapsulating node that has to pop the IOAM Indicator Label, data fields, and perform the IOAM function may not be capable of supporting it. The encapsulating node needs to know if the decapsulating node can support the IOAM function. The signaling extension for this capability exchange is outside the scope of this document.

The intermediate node that is not capable of supporting the IOAM functions defined in this document, can simply skip the IOAM processing of the MPLS header.

8. Data Packets with SR-MPLS Header

Segment Routing (SR) technology leverages the source routing paradigm [RFC8660]. A node steers a packet through a controlled set of instructions, called segments, by pre-pending the packet with an SR header. In the SR with MPLS data plane (SR-MPLS), the SR header is instantiated through a label stack.

An example of data packet carrying the SR-MPLS header with Path Segment Identifier (PSID) [I-D.ietf-spring-mpls-path-segment] with IOAM encapsulation is shown in Figure 2. The Path Segment Identifier allows to identify the path associated with the data traffic being monitored for IOAM on the decapsulating node.

```
  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
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                Label(1)               | TC  |S|      TTL      |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 .                                                               .
 .                                                               .
 .                                                               .
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                Label(n)               | TC  |S|      TTL      |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                PSID                   | TC  |S|      TTL      |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                Packet as shown in Figure 1                    |
 .                                                               .
 +---------------------------------------------------------------+
```

Figure 2: Data Packet with SR-MPLS Header

9. Security Considerations

The security considerations of SR-MPLS are discussed in [RFC8660], and the security considerations of IOAM in general are discussed in [I-D.ietf-ippm-ioam-data].

IOAM is considered a "per domain" feature, where one or several operators decide on leveraging and configuring IOAM according to their needs. Still, operators need to properly secure the IOAM domain to avoid malicious configuration and use, which could include injecting malicious IOAM packets into a domain.

10. IANA Considerations

IANA maintains the "Special-Purpose Multiprotocol Label Switching (MPLS) Label Values" registry (see <https://www.iana.org/assignments/mpls-label-values/mpls-label-values.xml>). IANA is requested to allocate IOAM Indicator Label value from the "Extended Special-Purpose MPLS Label Values" registry:

Value	Description	Reference
TBA1	E2E IOAM Indicator Label	This document
TBA2	HbH IOAM Indicator Label	This document

IANA maintains IP Version Number Registry (see <https://www.iana.org/assignments/version-numbers/version-numbers.xml>). IANA is requested to allocate IP Version Number 0010b for IOAM Data-type from "IP Version Numbers" registry.

11. References

11.1. Normative References

[I-D.ietf-ippm-ioam-data]
Brockners, F., Bhandari, S., and T. Mizrahi, "Data Fields for In-situ OAM", draft-ietf-ippm-ioam-data-11 (work in progress), November 2020.

[I-D.ietf-ippm-ioam-direct-export]
Song, H., Gafni, B., Zhou, T., Li, Z., Brockners, F., Bhandari, S., Sivakolundu, R., and T. Mizrahi, "In-situ OAM Direct Exporting", draft-ietf-ippm-ioam-direct-export-02 (work in progress), November 2020.

[I-D.ietf-ippm-ioam-flags]
Mizrahi, T., Brockners, F., Bhandari, S., Sivakolundu, R., Pignataro, C., Kfir, A., Gafni, B., Spiegel, M., and J. Lemon, "In-situ OAM Flags", draft-ietf-ippm-ioam-flags-03 (work in progress), October 2020.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

[RFC8660] Bashandy, A., Ed., Filsfils, C., Ed., Previdi, S., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing with the MPLS Data Plane", RFC 8660, DOI 10.17487/RFC8660, December 2019, <https://www.rfc-editor.org/info/rfc8660>.

[RFC8664] Sivabalan, S., Filsfils, C., Tantsura, J., Henderickx, W., and J. Hardwick, "Path Computation Element Communication Protocol (PCEP) Extensions for Segment Routing", RFC 8664, DOI 10.17487/RFC8664, December 2019, <https://www.rfc-editor.org/info/rfc8664>.

11.2. Informative References

[I-D.ietf-mpls-spl-terminology] Andersson, L., Kompella, K., and A. Farrel, "Special Purpose Label terminology", draft-ietf-mpls-spl-terminology-05 (work in progress), November 2020.

[I-D.ietf-spring-mpls-path-segment] Cheng, W., Li, H., Chen, M., Gandhi, R., and R. Zigler, "Path Segment in MPLS Based Segment Routing Network", draft-ietf-spring-mpls-path-segment-03 (work in progress), September 2020.

[RFC4385] Bryant, S., Swallow, G., Martini, L., and D. McPherson, "Pseudowire Emulation Edge-to-Edge (PWE3) Control Word for Use over an MPLS PSN", RFC 4385, DOI 10.17487/RFC4385, February 2006, <https://www.rfc-editor.org/info/rfc4385>.

[RFC4928] Swallow, G., Bryant, S., and L. Andersson, "Avoiding Equal Cost Multipath Treatment in MPLS Networks", BCP 128, RFC 4928, DOI 10.17487/RFC4928, June 2007, <https://www.rfc-editor.org/info/rfc4928>.

[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>.

Acknowledgements

The authors would like to thank Patrick Khordoc, Shwetha Bhandari and Vengada Prasad Govindan for the discussions on IOAM. The authors would also like to thank Tarek Saad, Loa Andersson, Greg Mirsky, and Cheng Li for providing many useful comments.

Contributors

Sagar Soni
Cisco Systems, Inc.

Email: sagsoni@cisco.com

Authors' Addresses

Rakesh Gandhi (editor)
Cisco Systems, Inc.
Canada

Email: rgandhi@cisco.com

Zafar Ali
Cisco Systems, Inc.

Email: zali@cisco.com

Clarence Filsfils
Cisco Systems, Inc.
Belgium

Email: cf@cisco.com

Frank Brockners
Cisco Systems, Inc.
Hansaallee 249, 3rd Floor
DUESSELDORF, NORDRHEIN-WESTFALEN 40549
Germany

Email: fbrockne@cisco.com

Bin Wen
Comcast

Email: Bin_Wen@cable.comcast.com

Voitek Kozak
Comcast

Email: Voitek_Kozak@comcast.com