IPPM Working Group
Internet-Draft
Intended status: Standards Track
Expires: March 16, 2019

X. Min
G. Mirsky
ZTE

September 12, 2018

Extended OAM to Carry In-situ OAM Configuration Data
draft-xiao-ippm-ioam-conf-state-01

Abstract

This document describes an extension for OAM packet such as IP Ping (ICMP [RFC0792] or ICMPv6 [RFC4443]) and MPLS LSP Ping [RFC8029], which can be used within an IOAM domain, allowing the IOAM encapsulating node to acquire IOAM configuration data of each IOAM transit node and/or IOAM decapsulating node easily and dynamically.

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 March 16, 2019.

Table of Contents

1. Introduction 2
  1.1. Conventions Used in This Document 3
    1.1.1. Terminology 3
    1.1.2. Requirements Language 3
2. IOAM Configuration Data Formats 3
  2.1. IOAM Configuration Data TLV 3
    2.1.1. IOAM Tracing Configuration Data sub-TLV 4
    2.1.2. IOAM Proof of Transit Configuration Data sub-TLV . . 6
    2.1.3. IOAM Edge-to-Edge Configuration Data sub-TLV 7
    2.1.4. IOAM End-of-Domain sub-TLV 8
3. Operational Guide 9
4. Security Considerations 9
5. IANA Considerations 9
6. Acknowledgements 10
7. Normative References 10
Authors' Addresses 11

1. Introduction

The Data Fields for In-situ OAM (IOAM) [I-D.ietf-ippm-ioam-data] defines data fields for IOAM which records OAM information within the packet while the packet traverses a particular network domain, which is called an IOAM domain. IOAM can be used to complement OAM mechanisms based on, e.g., ICMP or other types of probe packets, and IOAM mechanisms can be leveraged where mechanisms using, e.g., ICMP do not apply or do not offer the desired results.

As specified in [I-D.ietf-ippm-ioam-data], within the IOAM-domain, the IOAM data may be updated by network nodes that the packet traverses. The device which adds an IOAM data container to the packet to capture IOAM data is called the "IOAM encapsulating node", whereas the device which removes the IOAM data container is referred to as the "IOAM decapsulating node". Nodes within the domain which are aware of IOAM data and read and/or write or process the IOAM data are called "IOAM transit nodes". Both the IOAM encapsulating node and the decapsulating node are referred to as domain edge devices, which can be hosts or network devices.

In order to add accurate IOAM data container to the packet, the IOAM encapsulating node needs to know IOAM configurations at the IOAM transit nodes in a whole, e.g., how many IOAM transit nodes will add tracing data and what kinds of data fields will be added. Static configuration at the IOAM encapsulating node is a way to address this, but it's uneasy and inflexible, especially when the IOAM encapsulating node is a host. This document describes an extension for OAM packet such as IP Ping (ICMP [RFC0792] or ICMPv6 [RFC4443])

and MPLS LSP Ping [RFC8029], which can be used within an IOAM domain, allowing the IOAM encapsulating node to acquire IOAM configuration data of each IOAM transit node and/or IOAM decapsulating node easily and dynamically.

1.1. Conventions Used in This Document

1.1.1. Terminology

E2E: Edge to Edge

ICMP: Internet Control Message Protocol

IOAM: In-situ Operations, Administration, and Maintenance

LSP: Label Switched Path

MPLS: Multi-Protocol Label Switching

MTU: Maximum Transmission Unit

NTP: Network Time Protocol

OAM: Operations, Administration, and Maintenance

POSIX: Portable Operating System Interface

POT: Proof of Transit

PTP: Precision Time Protocol

TTL: Time to Live

1.1.2. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

2. IOAM Configuration Data Formats

2.1. IOAM Configuration Data TLV

IOAM Configuration Data uses TLV (Type-Length-Value tuple) which have the following format:

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Type = IOAM Configuration Data|            Length             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                             Value                             |
.                                                               .
.                                                               .
.                                                               .
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 1: IOAM Configuration Data TLV

When this TLV is present in the OAM packet sent by an IOAM encapsulating node, it means that the IOAM encapsulating node requests the receiving node to reply with its IOAM configuration data. If there is no IOAM configuration data to report by the receiving node, then this TLV SHOULD be ignored by the receiving node.

When this TLV is present in the OAM packet sent by an IOAM transit node and/or an IOAM decapsulating node, other than an IOAM encapsulating node, it means that IOAM function is enabled at this node and this TLV contains IOAM configuration data of the sender. Note that the IOAM encapsulating node or the IOAM decapsulating node can also be an IOAM transit node.

Type is set to the value (to be assigned by IANA) which indicates that it's an IOAM Configuration Data TLV.

Length is the length of the Value field in octets. When this TLV is present in the OAM packet sent by an IOAM encapsulating node, the Length field should be set to 0, and no Value field is included in the TLV. Otherwise, the Length field must not be set to 0.

Value is zero padded to align to a 4-octet boundary, and sub-TLVs MAY be contained in this field. Based on the data fields for IOAM specified in [I-D.ietf-ippm-ioam-data], four new sub-TLVs are defined in this document.

2.1.1. IOAM Tracing Configuration Data sub-TLV

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Sub-type = Tracing Conf Data  |            Length             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     IOAM-Trace-Type           |       Egress_if_MTU       |F|R|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Egress_if_id (short or wide format)         ......           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 2: IOAM Tracing Configuration Data Sub-TLV

When this sub-TLV is present in the IOAM Configuration Data TLV, it means that the sending node is an IOAM transit node and IOAM tracing function is enabled at this IOAM transit node.

Sub-type is set to the value (to be assigned by IANA) which indicates that it's an IOAM Tracing Configuration Data sub-TLV.

Length is the length of the sub-TLV's Value field in octets, if Egress_if_id in the short format which has 16 bits is used, it MUST be set to 6, and if Egress_if_id in the wide format which has 32 bits is used, it MUST be set to 8.

IOAM-Trace-Type field has the same name, length and definition as what's specified in both section 4.1.1 of [I-D.ietf-ippm-ioam-data].

Egress_if_MTU field has 14 bits and specifies the MTU of the egress interface out of which the sending node would forward the received OAM packet.

F bit is specified to indicate whether the pre-allocated trace or incremental trace is enabled. F bit is set to 1 when pre-allocated trace is enabled and set to 0 when the incremental trace is enabled. The meaning and difference of pre-allocated trace and incremental trace are described in section 4.1 of [I-D.ietf-ippm-ioam-data]. If the IOAM encapsulating node receives different F bit value from different IOAM transit node, then the IOAM encapsulating node will reserve data space in the IOAM header for the IOAM transit node that set F bit to 1, and the IOAM encapsulating node won't reserve data space in the IOAM header for the IOAM transit node that set F bit to 0.

R bit is reserved for future standardization.

Egress_if_id field has 16 bits (in short format) or 32 bits (in wide format) and specifies the identifier of the egress interface out of which the sending node would forward the received OAM packet.

2.1.2. IOAM Proof of Transit Configuration Data sub-TLV

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Sub-type = POT Conf Data    |            Length             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|IOAM-POT-Type|P|SoR|               Reserved                    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 3: IOAM Proof of Transit Configuration Data Sub-TLV

When this sub-TLV is present in the IOAM Configuration Data TLV, it means that the sending node is an IOAM transit node and IOAM proof of transit function is enabled at this IOAM transit node.

Sub-type is set to the value (to be assigned by IANA) which indicates that it's an IOAM Proof of Transit Configuration Data sub-TLV.

Length is the length of the sub-TLV's Value field in octets, and MUST be set to 4.

IOAM-POT-Type field and P bit have the same name, length and definition as what's specified in section 4.2 of [I-D.ietf-ippm-ioam-data]. If the IOAM encapsulating node receives IOAM-POT-Type and/or P bit values from an IOAM transit node that are different from its own, then the IOAM encapsulating node MAY choose to abandon the proof of transit function or to select one kind of IOAM-POT-Type and P bit, it's based on the policy applied to the IOAM encapsulating node.

SoR field has two bits which means the size of "Random" and "Cumulative" data, which are specified in section 4.2 of [I-D.ietf-ippm-ioam-data]. This document defines SoR as follow:

0b00 means 64-bit "Random" and 64-bit "Cumulative" data.

0b01~0b11: Reserved for future standardization

Reserved field is used for future standardization.

2.1.3. IOAM Edge-to-Edge Configuration Data sub-TLV

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Sub-type = E2E Conf Data    |            Length             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         IOAM-E2E-Type         |TSF|TSL|       Reserved        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 4: IOAM Edge to Edge Configuration Data Sub-TLV

When this sub-TLV is present in the IOAM Configuration Data TLV, it means that the sending node is an IOAM decapsulating node and IOAM edge to edge function is enabled at this IOAM decapsulating node. That is to say, if the IOAM encapsulating node receives this sub-TLV, the IOAM encapsulating node can determine that the node which sends this sub-TLV is an IOAM decapsulating node.

Sub-type is set to the value (to be assigned by IANA) which indicates that it's an IOAM Edge to Edge Configuration Data sub-TLV.

Length is the length of the sub-TLV's Value field in octets, and MUST be set to 4.

IOAM-E2E-Type field has the same name, length and definition as what's specified in section 4.3 of [I-D.ietf-ippm-ioam-data].

TSF field specifies the timestamp format used by the sending node. This document defines TSF as follow:

0b00: PTP timestamp format

0b01: NTP timestamp format

0b10: POSIX timestamp format

0b11: Reserved for future standardization

TSL field specifies the timestamp length used by the sending node. This document defines TSL as follow:

When TSF field is set to 0b00 which indicates PTP timestamp format:

0b00: 64-bit PTPv1 timestamp as defined in IEEE1588-2008 [IEEE1588v2]

0b01: 80-bit PTPv2 timestamp as defined in IEEE1588-2008 [IEEE1588v2]

0b10~0b11: Reserved for future standardization

When TSF field is set to 0b01 which indicates NTP timestamp format:

0b00: 32-bit NTP timestamp as defined in NTPv4 [RFC5905]

0b01: 64-bit NTP timestamp as defined in NTPv4 [RFC5905]

0b10: 128-bit NTP timestamp as defined in NTPv4 [RFC5905]

0b11: Reserved for future standardization

When TSF field is set to 0b10 or 0b11, the TSL field would be ignored.

Reserved field is used for future standardization.

2.1.4. IOAM End-of-Domain sub-TLV

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Sub-type = End of Domain    |            Length             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 5: IOAM End of Domain Sub-TLV

When this sub-TLV is present in the IOAM Configuration Data TLV, it means that the sending node is an IOAM decapsulating node. That is to say, if the IOAM encapsulating node receives this sub-TLV, the IOAM encapsulating node can determine that the node which sends this sub-TLV is an IOAM decapsulating node. When the IOAM Edge-to-Edge Configuration Data sub-TLV is present in the IOAM Configuration Data TLV sent by the IOAM decapsulating node, the IOAM End-of-Domain sub-TLV doesn't need to be present in the same IOAM Configuration Data TLV, otherwise the End-of-Domain sub-TLV MUST be present in the IOAM Configuration Data TLV sent by the IOAM decapsulating node. Since both the IOAM Edge-to-Edge Configuration Data sub-TLV and the IOAM End-of-Domain sub-TLV can be used to indicate that the sending node

is an IOAM decapsulating node, it's recommended to include only the IOAM Edge-to-Edge Configuration Data sub-TLV if IOAM edge to edge function is enabled at this IOAM decapsulating node.

Length is the length of the sub-TLV's Value field in octets, and MUST be set to 0.

3. Operational Guide

Once the IOAM encapsulating node is triggered to acquire IOAM configuration data of each IOAM transit node and/or IOAM decapsulating node, the IOAM encapsulating node will send a batch of OAM probe packets that include the IOAM Configuration Data TLV, first with TTL equal to 1 to reach the nearest node which may be an IOAM transit node or not, then with TTL equal to 2 to reach the second nearest node which also may be an IOAM transit node or not, on the analogy of this to increase 1 to TTL every time the IOAM encapsulating node sends a new OAM probe packet, until the IOAM encapsulating node receives OAM probe reply packet sent by the IOAM decapsulating node, which must contain the IOAM Configuration Data TLV including the IOAM Edge-to-Edge Configuration Data sub-TLV or the IOAM End-of-Domain sub-TLV.

The IOAM encapsulating node may be triggered by the device administrator, the network management, the network controller, or even the live user traffic, and the specific triggering mechanisms are outside the scope of this document.

Each IOAM transit node and/or IOAM decapsulating node that receives an OAM probe packet containing the IOAM Configuration Data TLV will send an OAM probe reply packet to the IOAM encapsulating node, and within the OAM probe reply packet, there must be an IOAM Configuration Data TLV containing one or more sub-TLVs. The IOAM Configuration Data TLV contained in the OAM probe packet will be ignored by the receiving node that is unaware of IOAM.

4. Security Considerations

Knowledge of the state of the IOAM domain may be considered confidential. Implementations SHOULD provide a means of filtering the addresses to which echo reply messages, ICMP/ICMPv6 or MPLS LSP Ping, may be sent.

5. IANA Considerations

To be added.

Editor's Note: For different OAM packet such as IP Ping (ICMP [RFC0792] or ICMPv6 [RFC4443]) and MPLS LSP Ping [RFC8029] different Type and Sub-type will be requested from IANA.

6. Acknowledgements

To be added.

7. Normative References

[I-D.ietf-ippm-ioam-data]
Brockners, F., Bhandari, S., Pignataro, C., Gredler, H., Leddy, J., Youell, S., Mizrahi, T., Mozes, D., Lapukhov, P., Chang, R., daniel.bernier@bell.ca, d., and J. Lemon, "Data Fields for In-situ OAM", draft-ietf-ippm-ioam-data-03 (work in progress), June 2018.

[IEEE1588v2]
Institute of Electrical and Electronics Engineers, "IEEE Std 1588-2008 - IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems", IEEE Std 1588-2008, 2008, <http://standards.ieee.org/findstds/standard/1588-2008.html>.

[RFC0792] Postel, J., "Internet Control Message Protocol", STD 5, RFC 792, DOI 10.17487/RFC0792, September 1981, <https://www.rfc-editor.org/info/rfc792>.

[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>.

[RFC4443] Conta, A., Deering, S., and M. Gupta, Ed., "Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification", STD 89, RFC 4443, DOI 10.17487/RFC4443, March 2006, <https://www.rfc-editor.org/info/rfc4443>.

[RFC5905] Mills, D., Martin, J., Ed., Burbank, J., and W. Kasch, "Network Time Protocol Version 4: Protocol and Algorithms Specification", RFC 5905, DOI 10.17487/RFC5905, June 2010, <https://www.rfc-editor.org/info/rfc5905>.

[RFC8029] Kompella, K., Swallow, G., Pignataro, C., Ed., Kumar, N., Aldrin, S., and M. Chen, "Detecting Multiprotocol Label Switched (MPLS) Data-Plane Failures", RFC 8029, DOI 10.17487/RFC8029, March 2017, <https://www.rfc-editor.org/info/rfc8029>.

[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>.

Authors' Addresses

Xiao Min
ZTE
Nanjing
China

Phone: +86 25 88016574
Email: xiao.min2@zte.com.cn

Greg Mirsky
ZTE
USA

Email: gregimirsky@gmail.com