

SPRING Working Group
Internet-Draft
Intended status: Standards Track
Expires: August 30, 2018

Z. Ali
C. Filsfils
R. Gandhi
N. Kumar
Cisco Systems, Inc.
D. Steinberg
Steinberg Consulting
S. Salsano
Universita di Roma "Tor Vergata"
P. L. Ventre
CNIT
G. Naik
Drexel University
D. Voyer
D. Bernier
Bell Canada
February 26, 2018

Performance Measurement in Segment Routing Networks with
IPv6 Data Plane (SRv6)
draft-ali-spring-srv6-pm-01.txt

Abstract

[RFC 6374](#) specifies protocol mechanisms to enable efficient and accurate measurement of packet loss, one-way and two-way delay, as well as related metrics such as delay variation and channel throughput in MPLS networks. This document describes how these mechanisms can be used for performance measurement of delay and loss in Segment Routing with IPv6 data plane (SRv6) networks.

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."

Internet-Draft

SRv6 Performance Measurement

February 26, 2018

Copyright Notice

Copyright (c) 2018 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1.	Introduction	3
2.	Conventions Used in This Document	3
2.1.	Key Word Definitions	3
2.2.	Abbreviations	3
2.3.	Terminology and Reference Topology	4
3.	Performance Delay Measurement	5
3.1.	One-Way Delay Measurement	5
3.2.	Two-Way Delay Measurement	6
3.3.	Delay Measurement Message Format	6
3.3.1.	Timestamps	8
3.4.	Delay Measurement Query Procedure	8
3.4.1.	Example Procedure	9
4.	Performance Loss Measurement	9
4.1.	One-Way Loss Measurement	9
4.2.	Two-Way Loss Measurement	10
4.3.	Loss Measurement Message Format	10
4.4.	Loss Measurement Query Procedure	12
4.4.1.	Example Procedure	12
5.	Probe Reply Message	13
5.1.	One-way Measurement Probe Reply	13
5.1.1.	Probe Reply Message to Controller	13
5.2.	Two-way Measurement Probe Reply	13
6.	Security Considerations	14
7.	IANA Considerations	14
8.	References	14
8.1.	Normative References	14

8.2. Informative References	14
Acknowledgments	16
Contributors	16
Authors' Addresses	16

[1.](#) Introduction

Service provider's requirements to satisfy Service Level Agreements (SLAs) depend on the ability to measure and monitor performance metrics for packet loss and one-way and two-way delay, as well as related metrics such as delay variation and channel throughput. The ability to monitor these performance metrics also provides operators with greater visibility into the performance characteristics of their networks, thereby facilitating planning, troubleshooting, and network performance evaluation.

[RFC6374] specifies protocol mechanisms to enable the efficient and accurate measurement of these performance metrics in MPLS networks. The One-Way Active Measurement Protocol (OWAMP) defined in [[RFC4656](#)] and Two-Way Active Measurement Protocol (TWAMP) defined in [[RFC5357](#)] provide capabilities for the measurement of various performance metrics in IP networks. However, mechanisms defined in [[RFC6374](#)] are more suitable for Segment Routing when using MPLS data plane (SR-MPLS) [[I-D.spring-sr-mpls-pm](#)]. This document describes how the mechanisms in [[RFC6374](#)] can be used for Performance Measurement (PM) of delay and loss in Segment Routing with the IPv6 data plane (SRv6) networks.

[2.](#) Conventions Used in This Document

[2.1.](#) Key Word Definitions

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.

[2.2.](#) Abbreviations

DM: Delay Measurement.

Figure 1: Reference Topology

In the reference topology:

Nodes N1, N2, and N4 are SRv6 capable nodes.

Nodes N3, N5 and N6 are classic IPv6 nodes.

Node 100 is a controller.

Node Nk has a classic IPv6 loopback address Bk::/128

Node Nk has Ak::/48 for its local SID space from which Local SIDs are explicitly allocated.

The IPv6 address of the nth Link between node X and Y at the X side is represented as 2001:DB8:X:Y:Xn::, e.g., the IPv6 address of link6 (the 2nd link) between N3 and N4 at N3 in Figure 1 is 2001:DB8:3:4:32::. Similarly, the IPv6 address of link5 (the 1st link between N3 and N4) at node N3 is 2001:DB8:3:4:31::.

Ali, et al.

Expires August 30, 2018

[Page 4]

Internet-Draft

SRv6 Performance Measurement

February 26, 2018

Ak::0 is explicitly allocated as the END function at Node k.

Ak::Cij is explicitly allocated as the END.X function at node k towards neighbor node i via jth Link between node i and node j. e.g., A2::C31 represents END.X at N2 towards N3 via link3 (the 1st link between N2 and N3). Similarly, A4::C52 represents the END.X at N4 towards N5 via link10.

<S1, S2, S3> represents a SID list where S1 is the first SID and S3 is the last SID. (S3, S2, S1; SL) represents the same SID list but encoded in the SRH format where the rightmost SID (S1) in the SRH is the first SID and the leftmost SID (S3) in the SRH is the last SID.

(SA, DA) (S3, S2, S1; SL) represents an IPv6 packet, SA is the IPv6 Source Address, DA the IPv6 Destination Address, (S3, S2, S1; SL) is the SRH header that includes the SID list <S1, S2, S3>.

SR policy is defined in Section 3 of [\[I-D.spring-segment-routing-policy\]](#).

[3.](#) Performance Delay Measurement

3.1. One-Way Delay Measurement

The one-way delay measurement for Packet IP network is defined in [RFC7679]. It is further exemplified using the following Figure 2.

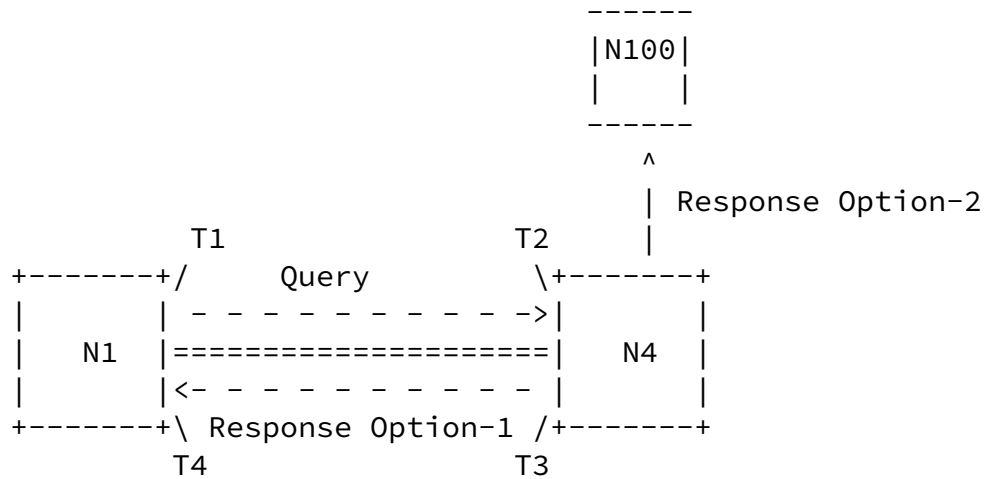


Figure 2: Delay Measurement Reference Model

Nodes N1 and N4 may not be directly connected, as shown in the reference topology in Figure 1. When N1 and N4 are not directly connected, the one-way delay measurement reflects the delay observed by the packet over an arbitrary SRv6 segment-list (SR policy)

[I-D.spring-segment-routing-policy]. In other words, the one-way delay is associated with the forward (N1 to N4) direction of the SRv6 segment-list.

In Figure 2, T1 refers to the time when the packet is transmitted from N1. Timestamp is added as late as possible at the egress pipeline (in hardware) at N1. T2 refers to the time when the packet is received at N4. Timestamp at the receiver (N4) is added as soon as possible at the ingress pipeline (in hardware).

The one-way delay metric can be computed as follow [RFC7679], [RFC6374]:

$$\text{One-way delay} = T2 - T1$$

3.2. Two-Way Delay Measurement

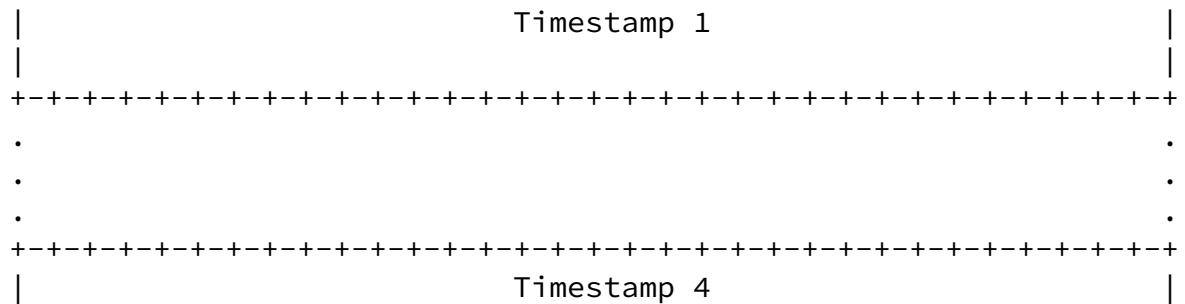
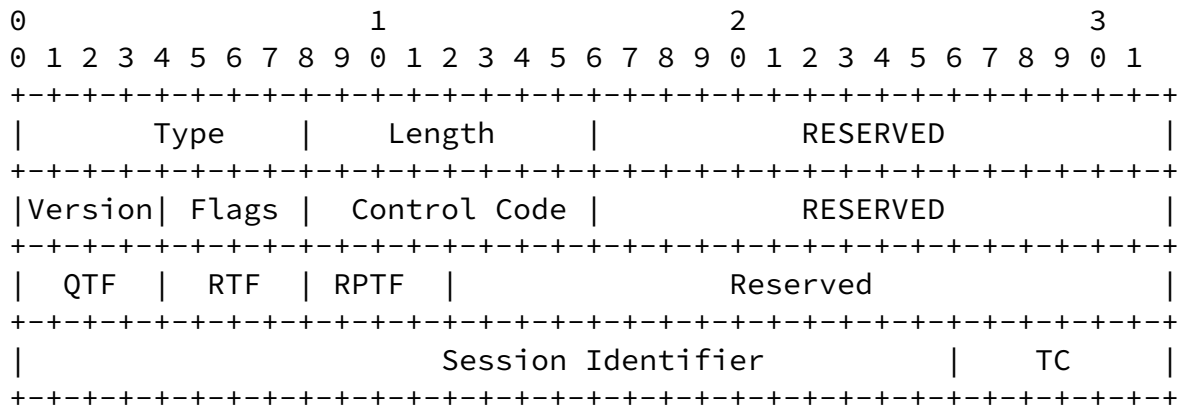
Similarly to the timestamps T1 and T2 in the forward direction, timestamps T3 and T4 are added in the DM packet in the reverse direction. T3 refers to the time the packet is transmitted from N4. T4 refers to the time the packet is received at N1.

The two-way delay metric can be computed as follows [[RFC6374](#)]:

$$\text{Two-way delay} = (T_4 - T_3) + (T_2 - T_1)$$

3.3. Delay Measurement Message Format

[I-D.6man-segment-routing-header] defines Segment Routing Header (SRH) for SRv6. SRH can contain TLVs, as specified in [I-D.6man-segment-routing-header]. This document defines Delay Measurement (DM) TLV that is carried in SRH for delay measurement. The DM TLV uses a modified DM message format specified in [RFC6374] and is defined as follows:



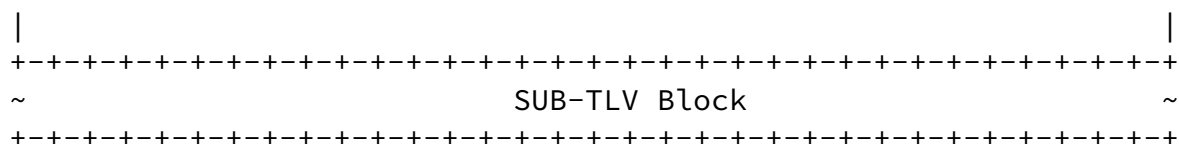


Figure 3: Delay Measurement TLV Format

The meanings of the fields are summarized in the following table.

Field	Meaning
Type	SRH DM TLV type (Value TBA by IANA)
Length	Total length of the TLV in bytes
Version	Protocol version
Flags	Message control flags
Control Code	Code identifying the query or response type
QTF	Querier timestamp format
RTF	Responder timestamp format
RPTF	Responder's preferred timestamp format
Reserved	Reserved for future specification
Session Identifier	Set arbitrarily by the querier
Traffic Class (TC) Field	Traffic Class being measured
Timestamp 1-4	64-bit timestamp values (see Section 3.4 in [RFC6374])
SUB-TLV Block	Optional block of Type-Length-Value fields

Reserved fields MUST be set to 0 and ignored upon receipt. The possible values for the remaining fields are as follows.

Version: Currently set to 1 (to identify definition of TC field in [\[RFC6374\]](#))

Flags: As specified in [\[RFC6374\]](#). The T flag in a DM message is set to 1.

Control Code: As specified in [\[RFC6374\]](#).

Message Length: Set to the total length of this message in bytes,

including the Version, Flags, Control Code, and Message Length fields

as well as the TLV Block, if any.

Querier Timestamp Format: The format of the timestamp values written by the querier, as specified in [Section 3.4 of \[RFC6374\]](#).

Responder Timestamp Format: The format of the timestamp values written by the responder, as specified in [Section 3.4 of \[RFC6374\]](#).

Responder's Preferred Timestamp Format: The timestamp format preferred by the responder, as specified in [Section 3.4 of \[RFC6374\]](#).

Session Identifier: Set arbitrarily in a query and copied in the response, if any. This field uniquely identifies a measurement operation (also called a session) that consists of a sequence of messages. All messages in the sequence have the same Session Identifier [\[RFC6374\]](#).

TC: Traffic Class being measured.

Timestamp 1-4 (T1-T4): The mapping of timestamps to the Timestamp 1-4 fields is designed to ensure that transmit timestamps are always written at the same fixed offset in the packet, and likewise for receive timestamps. This property is important for hardware processing.

SUB-TLV Block: Zero or more TLV fields. This document assumes the use of the DM message TLVs defined in [\[RFC6374\]](#).

[3.3.1.](#) Timestamps

[\[RFC6374\]](#), [Section 3.4](#) defines timestamp format that can be used for delay measurement. The IEEE 1588 Precision Time Protocol (PTP) timestamp format [\[IEEE1588\]](#) is used by default as described in [Appendix A of \[RFC6374\]](#), but it may require hardware support. As an alternative, Network Time Protocol (NTP) timestamp format is also supported in [\[RFC6374\]](#).

Note that for one-way delay measurement, Clock synchronization between the querier and responder nodes using methods detailed in [\[RFC6374\]](#) is required. Two-way delay measurement does not require clock to be synchronized between the querier and responder nodes.

[3.4.](#) Delay Measurement Query Procedure

For delay measurement using synthetic probes, a DM TLV is inserted in the SRH to record timestamps and SID function END.OTP as described in

3.4.1. Example Procedure

- o Node N1 constructs a DM probe packet with (B1::0, A2::C31)(A4::C52, A2::C31, SL=1; NH=NONE, DM TLV). To punt the DM probe packet at node N4, node N1 inserts the END.OTP SID [\[I-D.spring-srv6-network-programming\]](#) just before the target SID A4::C52 in the SRH. Hence, the packet as it leaves node N1 looks like (B1::0, A2::C31)(A4::C52, A4::OTP, A2::C31; SL=2; NH=NONE, DM TLV (with T1 from N1)). The PM synthetic probe query message does not contain any payload data.
- o When node N4 receives the packet (B1::0, A4::OTP)(A4::C52, A4::OTP, A2::C31; SL=1; NH=NONE, DM TLV), it processes the SID function END.OTP, as described in the pseudo code in [\[I-D.spring-srv6-network-programming\]](#). In doing so, it punts the timestamped packet (with T2 from N4) to the Performance Measurement (PM) process in control plane for processing.

4.1. One-Way Loss Measurement

```

      |N100|
      |    |
      -----
      ^
      | Response Option-2
      |
+-----+ /      Query      \ +-----+
|      N1      |=====|      N4      |
|      |      |<-----|      |      |
+-----+ \ Response Option-1 / +-----+

```

Figure 4: Loss Measurement Reference Model

Nodes N1 and N4 may not be directly connected, as shown in the reference topology in Figure 1. When N1 and N4 are not directly connected, the one-way loss measurement reflects the loss observed by the packet over an arbitrary SRv6 segment-list (SR policy) [[I-D.spring-segment-routing-policy](#)]. In other words, the one-way loss is associated with the forward (N1 to N4) direction of the SRv6 segment-list.

In Figure 4, C1 refers to the packet (or byte) count of traffic transmitted from N1. C2 refers to the packet (or byte) count of the traffic received at N4.

The one-way loss metric can be computed as follow [[RFC6374](#)]:

$$\text{One-way delay} = C2 - C1$$

[4.2.](#) Two-Way Loss Measurement

Similarly to the counters C1 and C2 in the forward direction, counters C3 and C4 are added in the LM packet in the reverse direction. C3 refers to the packet (or byte) count of traffic transmitted from N4. C4 refers to the packet (or byte) count of traffic received at N1.

The two-way loss metric can be computed as follows [[RFC6374](#)]:

$$\text{Two-way loss} = (C4 - C3) + (C2 - C1)$$

[4.3.](#) Loss Measurement Message Format

[I-D.6man-segment-routing-header] defines Segment Routing Header (SRH) for SRv6. SRH can contain TLVs, as specified in [[I-D.6man-segment-routing-header](#)]. This document defines Loss Measurement (LM) TLV for SRH. The LM TLV uses a modified LM message format specified in [[RFC6374](#)] and is defined as follows:

0

1

2

3

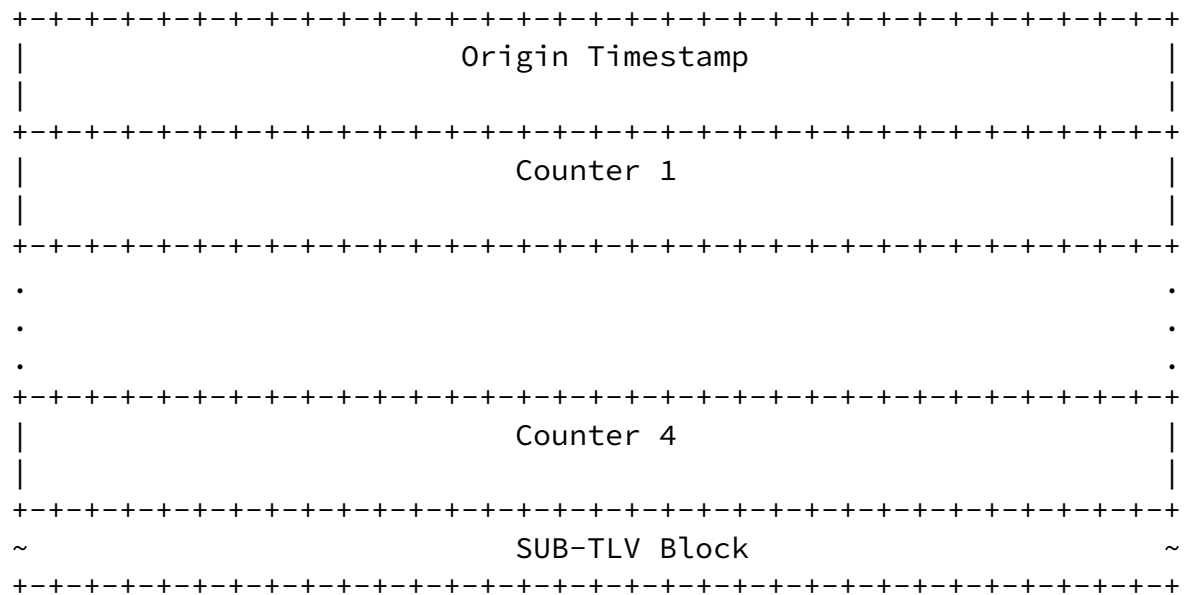


Figure 5: Loss Measurement TLV Format

The meanings of the fields are summarized in the following table.

Field	Meaning
Type	SRH LM TLV type (Value TBA by IANA)
Length	Total length of the TLV in bytes
Color (C)	Color flag of the Counters 1-4
Version	Protocol version
Flags	Message control flags
Control Code	Code identifying the query or response type
QTF	Querier timestamp format
RTF	Responder timestamp format
RPTF	Responder's preferred timestamp format

Reserved	Reserved for future specification
Session Identifier	Set arbitrarily by the querier
Traffic Class (TC) Field	Traffic Class being measured
Counters 1-4	64-bit counter values
SUB-TLV Block	Optional block of Type-Length-Value fields

Reserved fields MUST be set to 0 and ignored upon receipt. The possible values for the remaining fields are as follows.

Version: Currently set to 1 (to identify definition of TC field in [\[RFC6374\]](#))

Flags: As specified in [\[RFC6374\]](#). The T flag in a LM message is set to 1.

Control Code: As specified in [\[RFC6374\]](#).

Message Length: Set to the total length of this message in bytes, including the Version, Flags, Control Code, and Message Length fields as well as the TLV Block, if any.

Querier Timestamp Format: The format of the timestamp values written by the querier, as specified in [Section 3.4 of \[RFC6374\]](#).

Responder Timestamp Format: The format of the timestamp values written by the responder, as specified in [Section 3.4 of \[RFC6374\]](#).

Responder's Preferred Timestamp Format: The timestamp format preferred by the responder, as specified in [Section 3.4 of \[RFC6374\]](#).

Origin Timestamp: The timestamp value written by the querier, as specified in [Section 3.4 of \[RFC6374\]](#).

Session Identifier: Set arbitrarily in a query and copied in the response, if any. This field uniquely identifies a measurement operation (also called a session) that consists of a sequence of messages. All messages in the sequence have the same Session Identifier [\[RFC6374\]](#).

TC: Traffic Class being measured.

Counter 1-4 (C1-C4): 64-bit fields for LM counter values.

SUB-TLV Block: Zero or more TLV fields. This document assumes the use of the LM message TLVs defined in [[RFC6374](#)].

[4.4.](#) Loss Measurement Query Procedure

For loss measurement using synthetic probes, an LM TLV in the SRH is used to record packet (or byte) counters and SID function END.OTP as described in the pseudo code in [[I-D.spring-srv6-network-programming](#)] is used to punt probe packets.

[4.4.1.](#) Example Procedure

To measure packet loss from node N1 over an SRv6 Policy [[I-D.spring-segment-routing-policy](#)] that goes through a segment-list (A2::C31, A4::C52) to node N4, following procedure is defined for sending queries:

- o Node N1 constructs a LM probe packet with (B1::0, A2::C31)(A4::C52, A2::C31, SL=1; NH=NONE, LM TLV). To punt the LM probe packet at node N4, node N1 inserts the END.OTP SID

[[I-D.spring-srv6-network-programming](#)] just before the target SID A4::C52 in the SRH. Hence, the packet as it leaves node N1 looks like (B1::0, A2::C31)(A4::C52, A4::OTP, A2::C31; SL=2; NH=NONE, LM TLV (with C1 from N1)). The PM synthetic probe query message does not contain any payload data.

- o When node N4 receives the packet (B1::0, A4::OTP)(A4::C52, A4::OTP, A2::C31; SL=1; NH=NONE, LM TLV), it processes the END.OTP SID function, as described in the pseudo code in [[I-D.spring-srv6-network-programming](#)]. In doing so, it punts the packet (with C2 from N4) to the Performance Measurement (PM) process in control plane for processing.

[5.](#) Probe Reply Message

For one-way measurement, the receiver (node N4 in Figure 2) may send a response to the sender or to a controller (N100 in Figure 2). For

two-way measurement, the receiver sends a response to the sender.

[5.1.](#) One-way Measurement Probe Reply

For one-way performance measurement [[RFC7679](#)], the PM querier node can receive "out-of-bands" probe replies by properly setting the UDP Return Object (URO) TLV in the probe message. The URO TLV (Type=131) is defined in [[RFC7876](#)] and includes the UDP-Destination-Port and IP Address. In particular, if the querier sets its own IP address in the URO TLV the probe response is sent back by the responder node to the querier node.

The PM process in the control plane on the responder node copies the content of the DM TLV into the payload of the PM reply message.

[5.1.1.](#) Probe Reply Message to Controller

As shown in Figure 1, if the querier node N1 requires the probe reply to be sent to the controller N100, it sets the IP address of N100 in the Address field of the URO TLV of the PM probe query message.

The PM process in the control plane on the responder node copies the content of the DM TLV into the payload of the PM reply message.

[5.2.](#) Two-way Measurement Probe Reply

For two-way performance measurement [[RFC6374](#)], when using a bidirectional channel, the probe reply message is sent back to the querier node using a message similar to the probe query message as an SRv6 packet. In this case, the "control code" in the probe message

is set to "in-band response requested" [[RFC6374](#)].

[6.](#) Security Considerations

This document defines procedures for performance delay and loss measurement for SRv6 networks using the message formats defined in [[RFC6374](#)]. This document does not introduce any additional security considerations other than those covered in [[RFC6374](#)].

[7.](#) IANA Considerations

IANA is requested to allocate values for the new SRH TLV Types for Delay Measurement TLV and Loss Measurement TLV.

[8.](#) References

[8.1.](#) Normative References

- [IEEE1588] IEEE, "1588-2008 IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems", March 2008.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [RFC 2119](#), March 1997.
- [RFC6374] Frost, D. and S. Bryant, "Packet Loss and Delay Measurement for MPLS networks", [RFC 6374](#), September 2011.
- [RFC7876] Bryant, S., Sivabalan, S., and Soni, S., "UDP Return Path for Packet Loss and Delay Measurement for MPLS Networks", [RFC 7876](#), July 2016.
- [RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in [RFC 2119](#) Key Words", [RFC 8174](#), May 2017.
- [I-D.spring-srv6-network-programming] Filsfils, C. et al. "SRv6 Network Programming", [draft-filsfils-spring-srv6-network-programming](#), work in progress.
- [I-D.6man-segment-routing-header] Previdi, S., Filsfils, et al, "IPv6 Segment Routing Header (SRH)", [draft-ietf-6man-segment-routing-header](#), work in progress.

[8.2.](#) Informative References

Ali, et al. Expires August 30, 2018 [Page 14]

Internet-Draft SRv6 Performance Measurement February 26, 2018

- [RFC4656] Shalunov, S., Teitelbaum, B., Karp, A., Boote, J., and M. Zekauskas, "A One-way Active Measurement Protocol (OWAMP)", [RFC 4656](#), September 2006.

- [RFC5357] Hedayat, K., Krzanowski, R., Morton, A., Yum, K., and J. Babiarz, "A Two-Way Active Measurement Protocol (TWAMP)", [RFC 5357](#), October 2008.
- [RFC7679] Almes, G., et al, "A One-Way Delay Metric for IP Performance Metrics (IPPM)", [RFC 7679](#), January 2016.
- [I-D.spring-segment-routing-policy] Filsfils, C., et al. "Segment Routing Policy for Traffic Engineering", [draft-filsfils-spring-segment-routing-policy](#), work in progress.
- [I-D.spring-sr-mpls-pm] Filsfils, C., et al. "Segment Routing Policy for Traffic Engineering", [draft-gandhi-spring-sr-mpls-pm](#), work in progress.

Acknowledgments

To be added.

Contributors

Faisal Iqbal
Cisco Systems, Inc.
Email: faiqbal@cisco.com

Carlos Pignataro
Cisco Systems, Inc.
Email: cpignata@cisco.com

Authors' Addresses

Clarence Filsfils
Cisco Systems, Inc.
Email: cfilsfil@cisco.com

Zafar Ali
Cisco Systems, Inc.
Email: zali@cisco.com

Rakesh Gandhi
Cisco Systems, Inc.
Canada
Email: rgandhi@cisco.com

Nagendra Kumar
Cisco Systems, Inc.
Email: naikumar@cisco.com

Dirk Steinberg
Steinberg Consulting
Germany
Email: dws@dirksteinberg.de

Stefano Salsano
Universita di Roma "Tor Vergata"

Italy

Ali, et al.

Expires August 30, 2018

[Page 16]

Internet-Draft

SRv6 Performance Measurement

February 26, 2018

Email: stefano.salsano@uniroma2.it

Pier Luigi Ventre

CNIT

Italy

Email: pierluigi.ventre@cnit.it

Gaurav Naik

Drexel University

USA

Email: gn@drexel.edu

Daniel Voyer

Bell Canada

Email: daniel.voyer@bell.ca

Daniel Bernier

Bell Canada

Email: daniel.bernier@bell.ca

Ali, et al.

Expires August 30, 2018

[Page 17]