IPPM
Internet-Draft
Intended status: Standards Track
Expires: June 3, 2021

H. Song
D. Linda
Futurewei Technologies

November 30, 2020

In-band Edge-to-Edge Round Trip Time Measurement
draft-song-ippm-inband-e2e-rtt-measurement-00

Abstract

This draft describes a lightweight in-band edge-to-edge network round trip time measurement architecture and suggests implementations.

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.

Table of Contents

1. Introduction

In-network service-based traffic engineering or load balancing needs to monitor particular flows' edge-to-edge performance, such as round trip time (RTT), in the operator's network domain. The host-based ping using ICMPv6 [RFC4443] is of no use because it is usually beyond the access of network operators. The router-based ping, as an active measurement approach, cannot reflect the real performance of the specific flows under scrutiny. This is also true for the other active measurement approaches such as TWAMP [RFC5357].

In-situ OAM (IOAM) [I-D.ietf-ippm-ioam-data] supports in-band flow-based performance measurement. However, on the one hand, the IOAM trace option is too heavy for the applications which do not care about the per-hop performance; on the other hand, the IOAM E2E option only supports the one-way measurement.

Alternate Marking(AM) [RFC8321], mainly designed for one-way measurement, can be used to measure the two-way edge-to-edge delay if both edges initiate a one-way measurement session. However, AM's measurement interval needs to be large enough to avoid the measurement ambiguity, and it requires both edges to conduct the measurements and export results to a controller.

We need a lightweight in-band flow RTT measurement method. "Lightweight" means the extra header overhead is low, and the extra network processing overhead is also low. A network operator should be able to pick a flow to monitor and get find-grained per-packet RTT

measurement for edge to edge. Depending on the application scenario and the network domain scope, the edge can extend to the host, the network interface card (NIC), or the network switch or router. To this end, we propose an in-band edge-to-edge RTT measurement method and suggest the implementation approaches.

Such measurement only reflects the network delay for a flow but excludes the application layer delay incurred by server or client.

2. In-band E2E RTT Measurement Architecture

The measurement architecture is shown in Figure 1. The controller, either on a remote machine or on the edge node's control plane, configures the ingress edge node to measure some flow's RTT between the ingress edge and the egress edge. The ingress edge node uses ACL to filter the flow packets and, at given interval or probability, add the timestamp and the other metadata to the selected packets. The egress edge, after capturing the data, either piggyback the data on a reverse flow packet, or generate a feedback packet carrying the data back to the ingress edge node. Once the ingress edge node receives the feedback data, it sends the data along with the current timestamp to the controller. The controller can then calculate the flow RTT and react with followup actions.

The RTT calculation is done in the slow path, the metadata incurs only small and fix header overhead, and the nodes in the domain does not do any processing. All these make the measurement lightweight, accurate, and have little impact to the network forwarding performance.

```
                   +------+
                   |      |
                   |Ctrl. |
                   |      |
                   +-+----+
                     |  ^
             Config. |  | Export
                     V  |
        +------+      +----+-+    Forwarding    +------+      +------+
        |      | pkt. |      +-----............----->|      | pkt. |      |
        |Client+----->| Edge |                  | Edge +----->|Server|
        |      |      |      |<----.....-------+      |      |      |
        +------+      +------+    Feedback      +------+      +------+

                      |<--  Operator Network Domain -->|
```

Figure 1: In-band E2E RTT Measurement

To differentiate a feedback packet from an original packet, a flag needs to be raised in the feedback. Optionally, to correlate a feedback with its original packet, the original packet can also include an identifier (e.g., a sequence number) which the feedback packet will carry back as well. The ingress edge node can use the reverse flow ID plus the identifier to pair an original packet with its feedback.

The feedback can also include some other local data at the egress edge (e.g., the egress edge node ID or the egress flow statistics) other than simply reflecting the original data back.

3. Implementation Considerations

One approach to implement the in-band E2E RTT measurement is to use the IOAM E2E option augmented with the feedback mechanism. Current IOAM E2E option only sends one-way data from one edge to the other edge. The data fields can include the ingress edge timestamp which is exactly what is needed. Moreover, the data fields can also include a packet sequence number used for correlating the feedback packet with the original packet. However, current IOAM E2E option lacks a feedback mechanism. It has no flag field reserved in its current option header specification, so it is not easy to indicate the feedback packets.

To enable the two-way measurement behavior, we need to add some indicator to the IOAM E2E option header to indicate the request for a feedback. We also need another indicator to tell if the current packet is a feedback.

To support this, we can either introduce another IOAM two-way E2E option while keeping the current IOAM E2E option unchanged, or simply modify the current IOAM E2E option header specification to extend its usage. The simplest modification is to reserve a few flag bits and among them, two bits are used for the two-way measurement. One possible layout is shown in Figure 2. Alternatively, the flags can take several bits from the Namespace-ID field.

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|        Namespace-ID           | Flags |    IOAM-E2E-Type      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
|       E2E Option data field determined by IOAM-E2E-Type       |
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 2: Modified IOAM E2E Option Header

The data field can carry the timestamp, the sequence number, of a unique packet identifier number. Other data types can also be carried to enrich the feedback information.

A packet can serve as both a forward packet and a feedback packet when both flags are set. In this case, there are two records for each data type in the data field. The forward packet's data are located in front of the feedback packet's data.

4. Security Considerations

To prevent the timestamp to be maliciously altered during the packet forwarding, the ingress edge can instead keep the timestamp locally and only send a packet identifier (e.g., a random data). When a reverse flow packet carrying the same identifier is received, the current timestamp along with the saved timestamp are forwarded to the controller.

The ingress edge node can limit the frequency of measurement to the flow packets. The egress edge node can also rate limit the feedback. So the potential DoS attack can be mitigated.

5. IANA Considerations

Depending on the discussion output, either a registry for a new IOAM option is required or a modification to the current IOAM E2E option specification is needed.

6. Contributors

TBD.

7. Acknowledgments

TBD.

8. References

8.1. Normative References

[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>.

8.2. Informative 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.

[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>.

[RFC5357] Hedayat, K., Krzanowski, R., Morton, A., Yum, K., and J. Babiarz, "A Two-Way Active Measurement Protocol (TWAMP)", RFC 5357, DOI 10.17487/RFC5357, October 2008, <https://www.rfc-editor.org/info/rfc5357>.

[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

Haoyu Song
Futurewei Technologies
Santa Clara
USA

Email: haoyu.song@futurewei.com

Linda Dunbar
Futurewei Technologies
Plano
USA

Email: linda.dunbar@futurewei.com