IPPM
Internet-Draft
Intended status: Informational
Expires: August 22, 2021

H. Song
Futurewei Technologies
G. Mirsky
ZTE Corp.
C. Filsfils
A. Abdelsalam
Cisco Systems, Inc.
T. Zhou
Z. Li
Huawei
J. Shin
SK Telecom
K. Lee
LG U+

February 18, 2021

Postcard-based On-Path Flow Data Telemetry using Packet Marking
draft-song-ippm-postcard-based-telemetry-09

Abstract

The document describes a packet-marking variation of the Postcard-Based Telemetry (PBT), referred to as PBT-M. Unlike the instruction-based PBT, as embodied in IOAM DEX, PBT-M does not require the encapsulation of a telemetry instruction header, so it avoids some of the implementation challenges of the instruction-based PBT. However, PBT-M has unique issues that need to be considered. This document serves as a scheme overview and provides design guidelines applicable to implementations in different network protocols.

Status of This Memo

Copyright Notice

Table of Contents

1. Motivation

To gain detailed data plane visibility to support effective network OAM, it is essential to be able to examine the trace of user packets along their forwarding paths. Such on-path flow data reflect the state and status of each user packet's real-time experience and provide valuable information for network monitoring, measurement, and diagnosis.

The telemetry data include but not limited to the detailed forwarding path, the timestamp/latency at each network node, and, in case of packet drop, the drop location, and the reason. The emerging

programmable data plane devices allow user-defined data collection or conditional data collection based on trigger events. Such on-path flow data are from and about the live user traffic, which complements the data acquired through other passive and active OAM mechanisms such as IPFIX [RFC7011] and ICMP [RFC2925].

On-path telemetry was developed to cater to the need of collecting on-path flow data. There are two basic modes for on-path telemetry: the passport mode and the postcard mode. In the passport mode, each node on the path adds the telemetry data to the user packets (i.e., stamp the passport). The accumulated data-trace carried by user packets are exported at a configured end node. In the postcard mode, each node directly exports the telemetry data using an independent packet (i.e., send a postcard) to avoid the need for carrying the data with user packets.

In-situ OAM trace option (IOAM) [I-D.ietf-ippm-ioam-data] is a representative of the passport mode on-path telemetry. A prominent advantage of the passport mode is that it naturally retains the telemetry data correlation along the entire path. The passport mode also reduces the number of data export packets. These help to simplify the data collector and analyzer's work. On the other hand, the passport mode faces the following challenges.

o Issue 1: Since the telemetry instruction header and data processing must be done in the data-plane fast-path, it may interfere with the normal traffic forwarding (e.g., leading to forwarding performance degradation) and lead to inaccurate measurements (e.g., resulting in longer latency measurements than usual). This undesirable "observer effect" is problematic to carrier networks where stringent SLA must be observed.

o Issue 2: The passport mode may significantly increase the user packet's original size by adding data at each on-path node. The size may exceed the path MTU, so either the technique cannot apply, or the packet needs to be fragmented. That could be challenging when other network service headers (e.g., segment routing or service function chaining) are also present. Limiting the data size or path length reduces the effectiveness of INT.

o Issue 3: The instruction header needs to be encapsulated into user packets for transport. [I-D.brockners-inband-oam-transport] has discussed several encapsulation approaches for different transport protocols. So far, there is no feasible solution to encapsulate the instruction header in MPLS and IPv4 networks, which are still the most widely deployed. It is also challenging to encapsulate the instruction header in IPv6 [I-D.song-ippm-ioam-ipv6-support].

- Issue 4: The telemetry information is transported in plain text along the network paths. The instruction header and data are vulnerable to eavesdropping and tampering as well as DoS attack. Extra protective measurement is difficult on the data-plane fast-path.

- Issue 5: Since the passport mode only exports the telemetry data at the designated end node, if the packet is dropped in the network, the data will be lost as well. It cannot pinpoint the packet drop location, which is desired by fault diagnosis. Even worse, the end node may be unaware of the packet and data loss at all.

The postcard mode provides a perfect complement to the passport mode. In the variant of the postcard-based telemetry (PBT) which uses an instruction header, the postcards that carry telemetry data can be generated by a node's slow path and transported in-band or out-of-band, independent of the original user packets. IOAM direct export option (DEX) [I-D.ietf-ippm-ioam-direct-export] is a representative of PBT. Since an instruction header is still needed while successfully addressing issue 2 and 5 and partially addressing issue 1 and 4, this type of instruction-based PBT still cannot address issue 3.

This document describes another variation of the postcard mode on-path telemetry, the marking-based PBT (PBT-M). Unlike the instruction-based PBT, PBT-M does not require the encapsulation of a telemetry instruction header, so it avoids some of the implementation challenges of the instruction-based PBT. However, PBT-M has unique issues that need to be considered. This document discusses the challenges and their solutions of the marking-based PBT.

2. PBT-M: Marking-based PBT

As the name suggests, PBT-M only needs a marking-bit in the existing headers of user packets to trigger the telemetry data collection and export. The sketch of PBT-M is as follows. If on-path data need to be collected, the user packet is marked at the path head node. At each PBT-aware node, if the mark is detected, a postcard (i.e., the dedicated OAM packet triggered by a marked user packet) is generated and sent to a collector. The postcard contains the data requested by the management plane. The requested data are configured by the management plane. Once the collector receives all the postcards for a single user packet, it can infer the packet's forwarding path and analyze the data set. The path end node is configured to unmark the packets to its original format if necessary.

The overall architecture of PBT-M is depicted in Figure 1.

```
                         +------------+        +-----------+
                         | Network    |        | Telemetry |
                         | Management |(-------| Data      |
                         |            |        | Collector |
                         +-----:------+        +-----------+
                               :                     ^
                               :configurations       |postcards
                               :                     |(OAM pkts)
                ...............:.....................|........
                :             :               :      |       :
                :   +---------:---+-----------:---+--+-------:---+
                :   |         :   |           :   |          :   |
                V   |         V   |           V   |          V   |
             +------+-+     +-----+--+     +------+-+     +------+-+
   usr pkts  | Head    |     | Path    |     | Path    |     | End     |
        ====>| Node    |====>| Node    |====>| Node    |====>| Node    |===>
             |         |     | A       |     | B       |     |         |
             +--------+     +--------+     +--------+     +--------+
           mark usr pkts  gen postcards  gen postcards  gen postcards
           gen postcards                                unmark usr pkts
```

Figure 1: Architecture of PBT-M

PBT-M aims to address the issues listed above. It also introduces some new benefits. The advantages of PBT-M are summarized as follows.

- o 1: PBT-M avoids augmenting user packets with new headers and introducing new data plane protocols. The telemetry data collecting signaling remains in the data plane.

- o 2: PBT-M is extensible for collecting arbitrary new data to support possible future use cases. The data set to be collected can be configured through the management plane or control plane. Since there is no limitation on the types of data, any data other than those defined in [I-D.ietf-ippm-ioam-data] can also be collected. Since there is no size constraint anymore, it is free to use a more flexible data set template for data type definition.

- o 3: PBT-M avoids interfering with the normal forwarding and affecting the forwarding performance. Hence, the collected data are free to be transported independently through in-band or out-of-band channels. The data collecting, processing, assembly, encapsulation, and transport are, therefore, decoupled from the forwarding of the corresponding user packets and can be performed in data-plane slow-path if necessary.

- 4: For PBT-M, the types of data collected from each node can vary depending on application requirements and node capability. This is either impossible or very difficult to be supported by the passport mode in which the instruction header conveys data types collected per node.

- 5: PBT-M makes it easy to secure the collected data without exposing it to unnecessary entities. For example, both the configuration and the telemetry data can be encrypted before being transported, so passive eavesdropping and a man-in-the-middle attack can both be deterred.

- 6: Even if a user packet under inspection is dropped at some node in the network, the postcards collected from the preceding nodes are still valid and can be used to diagnose the packet drop location and reason.

3. New Challenges

Although PBT-M addresses the issues of the passport mode telemetry and the instruction-based PBT, it introduces a few new challenges.

- Challenge 1 (Packet Marking): A user packet needs to be marked to trigger the path-associated data collection. Since the PBT-M does not augment user packets with any new header fields, it needs to reserve or reuse bits from the existing header fields. This raises a similar issue as in the Alternate Marking Scheme [RFC8321]

- Challenge 2 (Configuration): Since the packet header will not carry OAM instructions anymore, the data plane devices need to be configured to know what data to collect. However, in general, the forwarding path of a flow packet (due to ECMP or dynamic routing) is unknown beforehand (note that there are some notable exceptions, such as segment routing). If the per-flow customized data collection is required, configuring the data set for each flow at all data plane devices might be expensive in terms of configuration load and data plane resources.

- Challenge 3 (Data Correlation): Due to the variable transport latency, the dedicated postcard packets for a single packet may arrive at the collector out of order or be dropped in networks for some reason. In order to infer the packet forwarding path, the collector needs some information from the postcard packets to identify the user packet affiliation and the order of path node traversal.

- Challenge 4 (Load Overhead): Since each postcard packet has its header, the overall network bandwidth overhead of PBT is higher than IOAM. A large number of postcards could add processing pressure on data collecting servers. That can be used as an attack vector for DoS.

4. PBT-M Design Considerations

To address the above challenges, we propose several design details of PBT-M.

4.1. Packet Marking

To trigger the path-associated data collection, usually, a single bit from some header field is sufficient. While no such bit is available, other packet-marking techniques are needed. We discuss several possible application scenarios.

- IPv4. Alternate Marking (AM) [RFC8321] is an IP flow performance measurement framework that also requires a single bit for packet coloring. The difference is that AM does in-network measurement while PBT-M only collects and exports data at network nodes (i.e., the data analysis is done at the collector rather than in the network nodes). AM suggests to use some reserved bit of the Flag field or some unused bit of the TOS field. Actually, AM can be considered a sub-case of PBT-M, so that the same bit can be used for PBT-M. The management plane is responsible for configuring the actual operation mode.

- SFC NSH. The OAM bit in the NSH header can be used to trigger the on-path data collection [I-D.ietf-sfc-nsh]. PBT does not add any other metadata to NSH.

- MPLS. Instead of choosing a header bit, we take advantage of the synonymous flow label [I-D.bryant-mpls-synonymous-flow-labels] approach to mark the packets. A synonymous flow label indicates the on-path data should be collected and forwarded through a postcard.

- SRv6: A flag bit in SRH can be reserved to trigger the on-path data collection [I-D.song-6man-srv6-pbt]. SRv6 OAM [I-D.ietf-6man-spring-srv6-oam] has adopted the O-bit in SRH flags as the marking bit to trigger the telemetry.

4.2. Flow Path Discovery

In case the path that a flow traverses is unknown in advance, all PBT-aware nodes should be configured to react to the marked packets by exporting some basic data, such as node ID and TTL before a data set template for that flow is configured. This way, the management plane can learn the flow path dynamically.

If the management plane wants to collect the on-path data for some flow, it configures the head node(s) with a probability or time interval for the flow packet marking. When the first marked packet is forwarded in the network, the PBT-aware nodes will export the basic data set to the collector. Hence, the flow path is identified. If other data types need to be collected, the management plane can further configure the data set's template to the target nodes on the flow's path. The PBT-aware nodes collect and export data accordingly if the packet is marked and a data set template is present.

If the flow path is changed for any reason, the new path can be quickly learned by the collector. Consequently, the management plane controller can be directed to configure the nodes on the new path. The outdated configuration can be automatically timed out or explicitly revoked by the management plane controller.

4.3. Packet Identity for Export Data Correlation

The collector needs to correlate all the postcard packets for a single user packet. Once this is done, the TTL (or the timestamp, if the network time is synchronized) can be used to infer the flow forwarding path. The key issue here is to correlate all the postcards for the same user packet.

The first possible approach includes the flow ID plus the user packet ID in the OAM packets. For example, the flow ID can be the 5-tuple IP header of the user traffic, and the user packet ID can be some unique information pertaining to a user packet (e.g., the sequence number of a TCP packet).

If the packet marking interval is large enough, the flow ID is enough to identify a user packet. As a result, it can be assumed that all the exported postcard packets for the same flow during a short time interval belong to the same user packet.

Alternatively, if the network is synchronized, then the flow ID plus the timestamp at each node can also infer the postcard affiliation. However, some errors may occur under some circumstances. For example, two consecutive user packets from the same flows are marked, but one exported postcard from a node is lost. It is difficult for

the collector to decide to which user packet the remaining postcard is related. In many cases, such a rare error has no catastrophic consequence. Therefore it is tolerable.

4.4. Control the Load

PBT-M should not be applied to all the packets all the time. It is better to be used in an interactive environment where the network telemetry applications dynamically decide which subset of traffic is under scrutiny. The network devices can limit the PBT rate through sampling and metering. The PBT packets can be distributed to different servers to balance the processing load.

It is important to understand that the total amount of data exported by PBT-M is identical to that of IOAM. The only extra overhead is the packet header of the postcards. In the case of IOAM, it carries the data from each node throughout the path to the end node before exporting the aggregated data. On the other hand, PBT-M directly exports local data. The overall network bandwidth impact depends on the network topology and scale, and PBT-M could be more bandwidth efficient.

5. Implementation Recommendation

5.1. Configuration

The head node's ACL should be configured to filter out the target flows for telemetry data collection. Optionally, a flow packet sampling rate or probability could be configured to monitor a subset of the flow packets.

The telemetry data set that should be exported by postcards at each path node could be configured using the data set templates specified, for example, in IPFIX [RFC7011]. In future revisions, we will provide more details.

The PBT-aware path nodes could be configured to respond or ignore the marked packets.

5.2. Postcard Format

The postcard should use the same data export format as that used by IOAM. [I-D.spiegel-ippm-ioam-rawexport] proposes a raw format that can be interpreted by IPFIX. In future revisions, we will provide more details.

5.3. Data Correlation

Enough information should be included to help the collector to correlate and order the postcards for a single user packet. Section 4.3 provides several possible means. The application scenario and network protocol are important factors to determine the means to use. In future revisions, we will provide details for representative applications.

6. Security Considerations

Several security issues need to be considered.

- Eavesdrop and tamper: the postcards can be encrypted and authenticated to avoid such security threats.

- DoS attack: PBT can be limited to a single administrative domain. The mark must be removed at the egress domain edge. The node can rate-limit the extra traffic incurred by postcards.

7. IANA Considerations

No requirement for IANA is identified.

8. Contributors

We thank Alfred Morton who provided valuable suggestions and comments helping improve this draft.

9. Acknowledgments

TBD.

10. Informative References

[I-D.brockners-inband-oam-transport]
Brockners, F., Bhandari, S., Govindan, V., Pignataro, C., Gredler, H., Leddy, J., Youell, S., Mizrahi, T., Mozes, D., Lapukhov, P., and R. Chang, "Encapsulations for In-situ OAM Data", draft-brockners-inband-oam-transport-05 (work in progress), July 2017.

[I-D.bryant-mpls-synonymous-flow-labels]
Bryant, S., Swallow, G., Sivabalan, S., Mirsky, G., Chen, M., and Z. Li, "RFC6374 Synonymous Flow Labels", draft-bryant-mpls-synonymous-flow-labels-01 (work in progress), July 2015.

[I-D.ietf-6man-spring-srv6-oam]
Ali, Z., Filsfils, C., Matsushima, S., Voyer, D., and M. Chen, "Operations, Administration, and Maintenance (OAM) in Segment Routing Networks with IPv6 Data plane (SRv6)", draft-ietf-6man-spring-srv6-oam-07 (work in progress), July 2020.

[I-D.ietf-ippm-ioam-data]
Brockners, F., Bhandari, S., and T. Mizrahi, "Data Fields for In-situ OAM", draft-ietf-ippm-ioam-data-10 (work in progress), July 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-00 (work in progress), February 2020.

[I-D.ietf-sfc-nsh]
Quinn, P., Elzur, U., and C. Pignataro, "Network Service Header (NSH)", draft-ietf-sfc-nsh-28 (work in progress), November 2017.

[I-D.song-6man-srv6-pbt]
Song, H., "Support Postcard-Based Telemetry for SRv6 OAM", draft-song-6man-srv6-pbt-01 (work in progress), October 2019.

[I-D.song-ippm-ioam-ipv6-support]
Song, H., Li, Z., and S. Peng, "Approaches on Supporting IOAM in IPv6", draft-song-ippm-ioam-ipv6-support-00 (work in progress), March 2020.

[I-D.spiegel-ippm-ioam-rawexport]
Spiegel, M., Brockners, F., Bhandari, S., and R. Sivakolundu, "In-situ OAM raw data export with IPFIX", draft-spiegel-ippm-ioam-rawexport-01 (work in progress), October 2018.

[RFC2925] White, K., "Definitions of Managed Objects for Remote Ping, Traceroute, and Lookup Operations", RFC 2925, DOI 10.17487/RFC2925, September 2000, <https://www.rfc-editor.org/info/rfc2925>.

[RFC7011] Claise, B., Ed., Trammell, B., Ed., and P. Aitken, "Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of Flow Information", STD 77, RFC 7011, DOI 10.17487/RFC7011, September 2013, <https://www.rfc-editor.org/info/rfc7011>.

[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
2330 Central Expressway
Santa Clara, 95050
USA

Email: hsong@futurewei.com

Greg Mirsky
ZTE Corp.

Email: gregimirsky@gmail.com

Clarence Filsfils
Cisco Systems, Inc.
Belgium

Email: cfilsfil@cisco.com

Ahmed Abdelsalam
Cisco Systems, Inc.
Italy

Email: ahabdels@cisco.com

Tianran Zhou
Huawei
156 Beiqing Road
Beijing, 100095
P.R. China

Email: zhoutianran@huawei.com

Zhenbin Li
Huawei
156 Beiqing Road
Beijing, 100095
P.R. China

Email: lizhenbin@huawei.com

Jongyoon Shin
SK Telecom
South Korea

Email: jongyoon.shin@sk.com

Kyungtae Lee
LG U+
South Korea

Email: coolee@lguplus.co.kr