IPPM
Internet-Draft
Intended status: Informational
Expires: October 15, 2020

H. Song, Ed.
Futurewei
T. Zhou
Z. Li
Huawei
J. Shin
SK Telecom
K. Lee
LG U+

April 13, 2020

Postcard-based On-Path Flow Data Telemetry
draft-song-ippm-postcard-based-telemetry-07

Abstract

The document describes a variation of the Postcard-Based Telemetry (PBT), the marking-based PBT. Unlike the instruction-based PBT, as embodied in [I-D.ietf-ippm-ioam-direct-export], the marking-based PBT does not require the encapsulation of a telemetry instruction header so it avoids some of the implementation challenges of the instruction-based PBT. This documents discuss the issues and solutions of the marking-based PBT.

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 October 15, 2020.

Table of Contents

1. Motivation

In order to gain detailed data plane visibility to support effective network OAM, it is important 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 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 complement the data acquired through other passive and active OAM mechanisms such as IPFIX [RFC7011] and ICMP [RFC2925].

On-path telemetry was developed to cater the need for 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 of 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.

- 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.

- 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 techniuqe cannot apply or the packet needs to be fragmented. This is especially troubling when some other network service headers (e.g., segment routing or service functoin chaining) are also present. Limiting the data size or path length reduces the effectiveness of INT.

- 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. However, There is no feasible solutions so far to encapsulate the instruction header in MPLS and IPv4 networks which are still the most widely deployed. It is also challenging to encapsulate the instruciton header in IPv6 [I-D.song-ippm-ioam-ipv6-support].

- Issue 4: 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 postcard-based telemetry (PBT), 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 the Issue 2 and 5 and partially addressing the Issue 1 and 4, this type of instruction-based PBT still cannot address the 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, the marking-based PBT does not require the encapsulation of a telemetry instruction header so it avoids some of the implementation challenges of the instruction-based PBT. This documents discuss the issues and 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. The user packet, if its on-path data need to be collected, 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 through data set templates (as in IPFIX [RFC7011]). 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 depict in Figure 1.

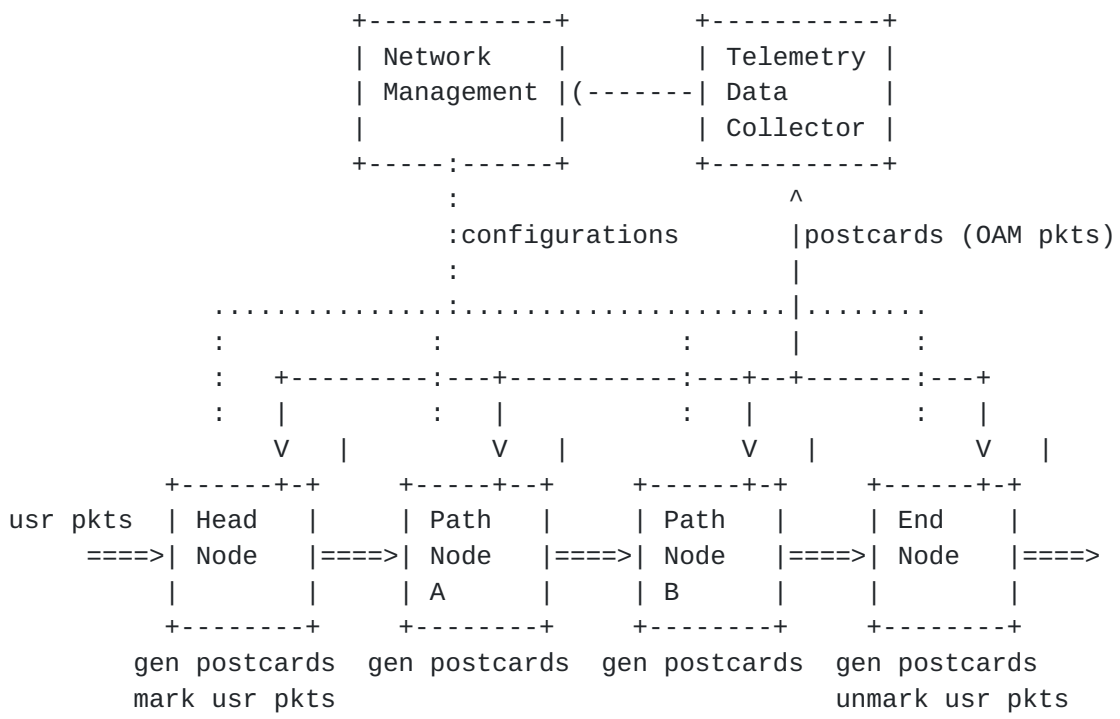

Figure 1: Architecture of PBT-M

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

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

- 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 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 constraints any more, it is free to use the more flexible data set template for data type definition.

- 3: PBT-M avoids interfering 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 data types collected per node are conveyed by the instruction header.

- 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 man-in-the-middle attack can both be deterred.

- 6: Even if a user packet under inspection is dropped at some node in network, the postcards that are collected from the previous 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: A user packet needs to be marked in order to trigger the path-associated data collection. Since we do not want to augment user packets with any new header fields, we must reuse some bit from existing header fields.

- Challenge 2: Since the packet header will not carry OAM instructions any more, 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). Configuring the data set for each flow at all data plane devices is expensive in terms of configuration load and data plane resources.

- Challenge 3: 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.

4. Considerations on PBT-M Design

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 three possible application scenarios.

- IPv4. IPFPM [I-D.ietf-ippm-alt-mark] is an IP flow performance measurement framework which also requires a single bit for packet coloring. The difference is that IPFPM 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). IPFPM suggests to use some reserved bit of the Flag field or some unused bit of the TOS field. Actually, IPFPM can be considered a subcase of PBT-M so the same bit can be used for PBT-M. The management plane is responsible to configure the actual operation mode.

- SFC NSH. The OAM bit in 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.

4.2. Flow Path Discovery

In case the path a flow traverses is unknown in advance, all PBT-aware nodes are 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 to the collector. Hence, the flow path is identified. If other types of data need to be collected, the management plane can further configure the data set template to the target nodes on the flow's path. The PBT-aware nodes would collect and export data

accordingly if the packet is marked and a data set template is present.

If for any reason the flow path is changed, the new path nodes can be learned immediately by the collector, so the management plane controller can be informed to configure the new path nodes. 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 a same user packet.

The first possible approach is to include 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, then the flow ID itself is enough to identify the user packet. That is, we can assume all the exported postcard packets for the same flow during a short period of time 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, if two consecutive user packets from the same flows are both marked but one exported postcard from a node is lost, then it is difficult for the collector to decide which user packet the remaining postcard belongs to. In many cases, such rare error has no catastrophic consequence therefore is tolerable.

4.4. Avoid Packet Marking through Node Configuration

It is possible to avoid needing to mark user packets yet still allowing in-band flow data collection. We could simply configure the Access Control List (ACL) to filter out the set of target flows. This approach has two potential issues: (1) Since the packet forwarding path is unknown in advance, one needs to configure all the nodes in a network to filter the flows and capture the complete data set. This wastes the precious ACL resource and is not scalable. (2) If a node cannot collect data for all the filtered packets of a flow, it needs to determine which packets to sample independently, so the

collector may not be able to receive the full set of postcards for a same user packet.

Nevertheless, since this approach does not require to touch the user packets at all, it has its unique merits: (1) User can freely choose any nodes as vantage points for data collection; (2) No need to worry that any "modified" user packets to leak out of the PBT domain; (3) It has the minimum impact to the forwarding of the user traffic.

No data plane standard is required to support this mode, except the postcard format.

5. Postcard Format

Postcard can 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.

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 administration 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

TBD.

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-ippm-alt-mark]
Fioccola, G., Capello, A., Cociglio, M., Castaldelli, L., Chen, M., Zheng, L., Mirsky, G., and T. Mizrahi, "Alternate Marking method for passive and hybrid performance monitoring", draft-ietf-ippm-alt-mark-14 (work in progress), December 2017.

[I-D.ietf-ippm-ioam-data]
Brockners, F., Bhandari, S., Pignataro, C., Gredler, H., Leddy, J., Youell, S., Mizrahi, T., Mozes, D., Lapukhov, P., remy@barefootnetworks.com, r., daniel.bernier@bell.ca, d., and J. Lemon, "Data Fields for In-situ OAM", draft-ietf-ippm-ioam-data-09 (work in progress), March 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-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>.

Authors' Addresses

Haoyu Song (editor)
Futurewei
2330 Central Expressway
Santa Clara, 95050
USA

Email: hsong@futurewei.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