IPPM
Internet-Draft
Intended status: Standards Track
Expires: January 13, 2022

H. Song
Futurewei
B. Gafni
Nvidia
T. Zhou
Z. Li
Huawei
F. Brockners
Cisco
S. Bhandari, Ed.
Thoughtspot
R. Sivakolundu
Cisco
T. Mizrahi, Ed.
Huawei

July 12, 2021

In-situ OAM Direct Exporting
draft-ietf-ippm-ioam-direct-export-05

Abstract

In-situ Operations, Administration, and Maintenance (IOAM) is used for recording and collecting operational and telemetry information. Specifically, IOAM allows telemetry data to be pushed into data packets while they traverse the network. This document introduces a new IOAM option type called the Direct Export (DEX) option, which is used as a trigger for IOAM data to be directly exported or locally aggregated without being pushed into in-flight data packets. The exporting method and format are outside the scope of this document.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction 2
2. Conventions 3
2.1. Requirement Language 3
2.2. Terminology 3
3. The Direct Exporting (DEX) IOAM Option Type 3
3.1. Overview 3
3.1.1. DEX Packet Selection 5
3.1.2. Responding to the DEX Trigger 5
3.2. The DEX Option Format 6
4. IANA Considerations 7
4.1. IOAM Type 7
4.2. IOAM DEX Flags 8
5. Performance Considerations 8
6. Security Considerations 8
7. References 9
7.1. Normative References 9
7.2. Informative References 10
Appendix A. Hop Limit and Hop Count in Direct Exporting 10
Authors' Addresses 11

1. Introduction

IOAM [I-D.ietf-ippm-ioam-data] is used for monitoring traffic in the network, and for incorporating IOAM data fields into in-flight data packets.

IOAM makes use of four possible IOAM options, defined in [I-D.ietf-ippm-ioam-data]: Pre-allocated Trace Option, Incremental Trace Option, Proof of Transit (POT) Option, and Edge-to-Edge Option.

This document defines a new IOAM option type (also known as an IOAM type) called the Direct Export (DEX) option. This option is used as a trigger for IOAM nodes to locally aggregate and process IOAM data, and/or to export it to a receiving entity (or entities). A "receiving entity" in this context can be, for example, an external collector, analyzer, controller, decapsulating node, or a software module in one of the IOAM nodes.

Note that even though the IOAM Option-Type is called "Direct Export", it depends on the deployment whether the receipt of a packet with DEX option type leads to the creation of another packet. Some deployments might simply use the packet with the DEX option type to trigger local processing of OAM data.

This draft has evolved from combining some of the concepts of PBT-I from [I-D.song-ippm-postcard-based-telemetry] with immediate exporting from [I-D.ietf-ippm-ioam-flags].

2. Conventions

2.1. Requirement 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.

2.2. Terminology

Abbreviations used in this document:

IOAM: In-situ Operations, Administration, and Maintenance

OAM: Operations, Administration, and Maintenance

DEX: Direct EXporting

3. The Direct Exporting (DEX) IOAM Option Type

3.1. Overview

The DEX option is used as a trigger for collecting IOAM data locally or for exporting it to a receiving entity (or entities). Specifically, the DEX option can be used as a trigger for collecting IOAM data by an IOAM node and locally aggregating it; thus, this aggregated data can be periodically pushed to a receiving entity, or pulled by a receiving entity on-demand.

This option is incorporated into data packets by an IOAM encapsulating node, and removed by an IOAM decapsulating node, as illustrated in Figure 1. The option can be read but not modified by transit nodes. Note: the terms IOAM encapsulating, decapsulating and transit nodes are as defined in [I-D.ietf-ippm-ioam-data].

```
                                  ^
                                  |Exported IOAM data
                                  |
                                  |
                                  |
              +--------------+------+-------+--------------+
              |              |              |              |
              |              |              |              |
User      +---+----+     +---+----+     +---+----+     +---+----+
packets   |Encapsu-|     | Transit|     | Transit|     |Decapsu-|
--------->|lating  |====>| Node   |====>| Node   |====>|lating  |---->
          |Node    |     | A      |     | B      |     |Node    |
          +--------+     +--------+     +--------+     +--------+
          Insert DEX       Export         Export       Remove DEX
          option and      IOAM data      IOAM data     option and
          export data                                  export data
```

Figure 1: DEX Architecture

The DEX option is used as a trigger to collect and/or export IOAM data. The trigger applies to transit nodes, the decapsulating node, and the encapsulating node:

o An IOAM encapsulating node configured to incorporate the DEX option encapsulates (possibly a subset of) the packets it forwards with the DEX option, and MAY export and/or collect the requested IOAM data immediately. Only IOAM encapsulating nodes are allowed to add the DEX option type to a packet.

o A transit node that processes a packet with the DEX option MAY export and/or collect the requested IOAM data.

o An IOAM decapsulating node that processes a packet with the DEX option MAY export and/or collect the requested IOAM data, and MUST decapsulate the IOAM header.

As in [I-D.ietf-ippm-ioam-data], the DEX option can be incorporated into all or a subset of the traffic that is forwarded by the encapsulating node, as further discussed in Section 3.1.1 below. Moreover, IOAM nodes respond to the DEX trigger by exporting and/or

collection IOAM data either for all traversing packets that carry the DEX option, or selectively only for a subset of these packets, as further discussed in Section 3.1.2 below.

3.1.1. DEX Packet Selection

If an IOAM encapsulating node incorporates the DEX option into all the traffic it forwards it may lead to an excessive amount of exported data, which may overload the network and the receiving entity. Therefore, an IOAM encapsulating node that supports the DEX option MUST support the ability to incorporate the DEX option selectively into a subset of the packets that are forwarded by it.

Various methods of packet selection and sampling have been previously defined, such as [RFC7014] and [RFC5475]. Similar techniques can be applied by an IOAM encapsulating node to apply DEX to a subset of the forwarded traffic.

The subset of traffic that is forwarded or transmitted with a DEX option SHOULD NOT exceed 1/N of the interface capacity on any of the IOAM encapsulating node's interfaces. It is noted that this requirement applies to the total traffic that incorporates a DEX option, including traffic that is forwarded by the IOAM encapsulating node and probe packets that are generated by the IOAM encapsulating node. In this context N is a parameter that can be configurable by network operators. If there is an upper bound, M, on the number of IOAM transit nodes in any path in the network, then it is recommended to use an N such that N >> M. The rationale is that a packet that includes a DEX option may trigger an exported packet from each IOAM transit node along the path for a total of M exported packets. Thus, if N >> M then the number of exported packets is significantly lower than the number of data packets forwarded by the IOAM encapsulating node. If there is no prior knowledge about the network topology or size, it is recommended to use N>100.

3.1.2. Responding to the DEX Trigger

The DEX option specifies which data fields should be exported and/or collected, as specified in Section 3.2. As mentioned above, the data can be locally collected, and optionally can be aggregated and exported to a receiving entity, either proactively or on-demand. If IOAM data is exported, the format and encapsulation of the packet that contains the exported data is not within the scope of the current document. For example, the export format can be based on [I-D.spiegel-ippm-ioam-rawexport].

An IOAM node that performs DEX-triggered exporting MUST support the ability to limit the rate of the exported packets. The rate of

exported packets SHOULD be limited so that the number of exported packets is significantly lower than the number of packets that are forwarded by the device. The exported data rate SHOULD NOT exceed 1/N of the interface capacity on any of the IOAM node's interfaces. It is recommended to use N>100. Depending on the IOAM node's architecture considerations, the export rate may be limited to a lower number in order to avoid loading the IOAM node.

Exported packets SHOULD NOT be exported over a path or a tunnel that is subject to IOAM direct exporting. Furthermore, IOAM encapsulating nodes that can identify a packet as an IOAM exported packet MUST NOT push a DEX option into such a packet. This requirement is intended to prevent nested exporting and/or exporting loops.

A transit IOAM node that does not support the DEX option SHOULD ignore it. A decapsulating node that does not support the DEX option MUST remove it, along with any other IOAM options carried in the packet if such exist.

3.2. The DEX Option Format

The format of the DEX option is depicted in Figure 2. The length of the DEX option is either 8 octets or 16 octets, as the Flow ID and the Sequence Number fields (summing up to 8 octets) are optional. It is assumed that the lower layer protocol indicates the length of the DEX option, thus indicating whether the two optional fields are present.

```
 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-Trace-Type                 |   Reserved    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                         Flow ID (optional)                    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Sequence Number  (Optional)               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 2: DEX Option Format

Namespace-ID   A 16-bit identifier of the IOAM namespace, as defined in [I-D.ietf-ippm-ioam-data].

Flags A 16-bit field, comprised of 16 one-bit subfields. Flags are allocated by IANA, as defined in Section 4.2.

IOAM-Trace-Type A 24-bit identifier which specifies which data fields should be exported. The format of this field is as defined in [I-D.ietf-ippm-ioam-data]. Specifically, bit 23, which corresponds to the Checksum Complement data field, should be assigned to be zero by the IOAM encapsulating node, and ignored by transit and decapsulating nodes. The reason for this is that the Checksum Complement is intended for in-flight packet modifications and is not relevant for direct exporting.

Reserved This field SHOULD be ignored by the receiver.

Flow ID A 32-bit flow identifier. If the actual Flow ID is shorter than 32 bits, it is zero padded in its most significant bits. The field is set at the encapsulating node. The Flow ID can be uniformly assigned by a central controller or algorithmically generated by the encapsulating node. The latter approach cannot guarantee the uniqueness of Flow ID, yet the conflict probability is small due to the large Flow ID space. The Flow ID can be used to correlate the exported data of the same flow from multiple nodes and from multiple packets.

Sequence Number A 32-bit sequence number starting from 0 and increasing by 1 for each following monitored packet from the same flow at the encapsulating node. The Sequence Number, when combined with the Flow ID, provides a convenient approach to correlate the exported data from the same user packet.

4. IANA Considerations

4.1. IOAM Type

The "IOAM Type Registry" was defined in Section 7.2 of [I-D.ietf-ippm-ioam-data]. IANA is requested to allocate the following code point from the "IOAM Type Registry" as follows:

TBD-type IOAM Direct Export (DEX) Option Type

If possible, IANA is requested to allocate code point 4 (TBD-type).

4.2. IOAM DEX Flags

IANA is requested to define an "IOAM DEX Flags" registry. This registry includes 16 flag bits. Allocation should be performed based on the "RFC Required" procedure, as defined in [RFC8126].

5. Performance Considerations

The DEX option triggers IOAM data to be collected and/or exported packets to be exported to a receiving entity (or entities). In some cases this may impact the receiving entity's performance, or the performance along the paths leading to it.

Therefore, the performance impact of these exported packets is limited by taking two measures: at the encapsulating nodes, by selective DEX encapsulation (Section 3.1.1), and at the transit nodes, by limiting exporting rate (Section 3.1.2). These two measures ensure that direct exporting is used at a rate that does not significantly affect the network bandwidth, and does not overload the receiving entity. Moreover, it is possible to load balance the exported data among multiple receiving entities, although the exporting method is not within the scope of this document.

6. Security Considerations

The security considerations of IOAM in general are discussed in [I-D.ietf-ippm-ioam-data]. Specifically, an attacker may try to use the functionality that is defined in this document to attack the network.

An attacker may attempt to overload network devices by injecting synthetic packets that include the DEX option. Similarly, an on-path attacker may maliciously incorporate the DEX option into transit packets, or maliciously remove it from packets in which it is incorporated.

Forcing DEX, either in synthetic packets or in transit packets may overload the receiving entity (or entities). Since this mechanism affects multiple devices along the network path, it potentially amplifies the effect on the network bandwidth and on the receiving entity's load.

The amplification effect of DEX may be worse in wide area networks in which there are multiple IOAM domains. For example, if DEX is used in IOAM domain 1 for exporting IOAM data to a receiving entity, then the exported packets of domain 1 can be forwarded through IOAM domain 2, in which they are subject to DEX. The exported packets of domain 2 may in turn be forwarded through another IOAM domain (or through

domain 1), and theoretically this recursive amplification may continue infinitely.

In order to mitigate the attacks described above, the following requirements (Section 3) have been defined:

- Selective DEX (Section 3.1.1) is applied by IOAM encpsulating nodes in order to limit the potential impact of DEX attacks to a small fraction of the traffic.

- Rate limiting of exported traffic (Section 3.1.2) is applied by IOAM nodes in order to prevent overloading attacks and in order to significantly limit the scale of amplification attacks.

- IOAM encapsulating nodes are required to avoid pushing the DEX option into IOAM exported packets (Section 3.1.2), thus preventing some of the amplification and export loop scenarios.

Although the exporting method is not within the scope of this document, any exporting method MUST secure the exported data from the IOAM node to the receiving entity. Specifically, an IOAM node that performs DEX exporting MUST send the exported data to a pre-configured trusted receiving entity.

IOAM is assumed to be deployed in a restricted administrative domain, thus limiting the scope of the threats above and their affect. This is a fundamental assumption with respect to the security aspects of IOAM, as further discussed in [I-D.ietf-ippm-ioam-data].

7. References

7.1. Normative 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-12 (work in progress), February 2021.

[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>.

[RFC5475] Zseby, T., Molina, M., Duffield, N., Niccolini, S., and F. Raspall, "Sampling and Filtering Techniques for IP Packet Selection", RFC 5475, DOI 10.17487/RFC5475, March 2009, <https://www.rfc-editor.org/info/rfc5475>.

[RFC7014] D'Antonio, S., Zseby, T., Henke, C., and L. Peluso, "Flow Selection Techniques", RFC 7014, DOI 10.17487/RFC7014, September 2013, <https://www.rfc-editor.org/info/rfc7014>.

[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>.

7.2. Informative References

[I-D.ietf-ippm-ioam-flags]
Mizrahi, T., Brockners, F., Bhandari, S., Sivakolundu, R., Pignataro, C., Kfir, A., Gafni, B., Spiegel, M., and J. Lemon, "In-situ OAM Flags", draft-ietf-ippm-ioam-flags-04 (work in progress), February 2021.

[I-D.song-ippm-postcard-based-telemetry]
Song, H., Mirsky, G., Filsfils, C., Abdelsalam, A., Zhou, T., Li, Z., Shin, J., and K. Lee, "Postcard-based On-Path Flow Data Telemetry using Packet Marking", draft-song-ippm-postcard-based-telemetry-09 (work in progress), February 2021.

[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-04 (work in progress), November 2020.

[RFC8126] Cotton, M., Leiba, B., and T. Narten, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 8126, DOI 10.17487/RFC8126, June 2017, <https://www.rfc-editor.org/info/rfc8126>.

Appendix A. Hop Limit and Hop Count in Direct Exporting

In order to help correlate and order the exported packets, it is possible to include the Hop_Lim/Node_ID data field in exported packets; if the IOAM-Trace-Type [I-D.ietf-ippm-ioam-data] has the Hop_Lim/Node_ID bit set, then exported packets include the Hop_Lim/Node_ID data field, which contains the TTL/Hop Limit value from a lower layer protocol.

An alternative approach was considered during the design of this document, according to which a 1-octet Hop Count field would be included in the DEX header (presumably by claiming some space from the Flags field). The Hop Limit would starts from 0 at the encapsulating node and be incremented by each IOAM transit node that

supports the DEX option. In this approach the Hop Count field value would also be included in the exported packet.

The main advantage of the Hop_Lim/Node_ID approach is that it provides information about the current hop count without requiring each transit node to modify the DEX option, thus simplifying the data plane functionality of Direct Exporting. The main advantage of the Hop Count approach that was considered is that it counts the number of IOAM-capable nodes without relying on the lower layer TTL, especially when the lower layer cannot prvide the accurate TTL information, e.g., Layer 2 Ethernet or hierarchical VPN. The Hop Count approach would also explicitly allow to detect a case where an IOAM-capable node fails to export packets. It would also be possible to use a flag to indicate an optional Hop Count field, which enables to control the tradeoff. On one hand it would address the use cases that the Hop_Lim/Node_ID cannot cover, and on the other hand it would not require transit switches to update the option if it was not supported or disabled. For the sake of simplicity the Hop Count approach was not pursued, and this field is not incorporated in the DEX header.

Authors' Addresses

Haoyu Song
Futurewei
2330 Central Expressway
Santa Clara 95050
USA

Email: haoyu.song@futurewei.com

Barak Gafni
Nvidia
350 Oakmead Parkway, Suite 100
Sunnyvale, CA 94085
U.S.A.

Email: gbarak@nvidia.com

Tianran Zhou
Huawei
156 Beiqing Rd.
Beijing 100095
China

Email: zhoutianran@huawei.com

Zhenbin Li
Huawei
156 Beiqing Rd.
Beijing 100095
China

Email: lizhenbin@huawei.com

Frank Brockners
Cisco Systems, Inc.
Hansaallee 249, 3rd Floor
DUESSELDORF, NORDRHEIN-WESTFALEN 40549
Germany

Email: fbrockne@cisco.com

Shwetha Bhandari (editor)
Thoughtspot
3rd Floor, Indiqube Orion, 24th Main Rd, Garden Layout, HSR Layout
Bangalore, KARNATAKA 560 102
India

Email: shwetha.bhandari@thoughtspot.com

Ramesh Sivakolundu
Cisco Systems, Inc.
170 West Tasman Dr.
SAN JOSE, CA 95134
U.S.A.

Email: sramesh@cisco.com

Tal Mizrahi (editor)
Huawei
8-2 Matam
Haifa 3190501
Israel

Email: tal.mizrahi.phd@gmail.com