IPPM                                                        H. Song
Internet-Draft                                            Futurewei
Intended status: Standards Track                          B. Gafni
Expires: August 21, 2021                                    Nvidia
                                                           T. Zhou
                                                             Z. Li
                                                           Huawei
                                                      F. Brockners
                                                             Cisco
                                                  S. Bhandari, Ed.
                                                       Thoughtspot
                                                   R. Sivakolundu
                                                             Cisco
                                                   T. Mizrahi, Ed.
                                                           Huawei
                                                 February 17, 2021

                       In-situ OAM Direct Exporting
                    draft-ietf-ippm-ioam-direct-export-03

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 without being
   pushed into in-flight data packets.

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 August 21, 2021.

Table of Contents

## [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 export IOAM data 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.

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 exporting telemetry data to a
receiving entity (or entities).

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
```
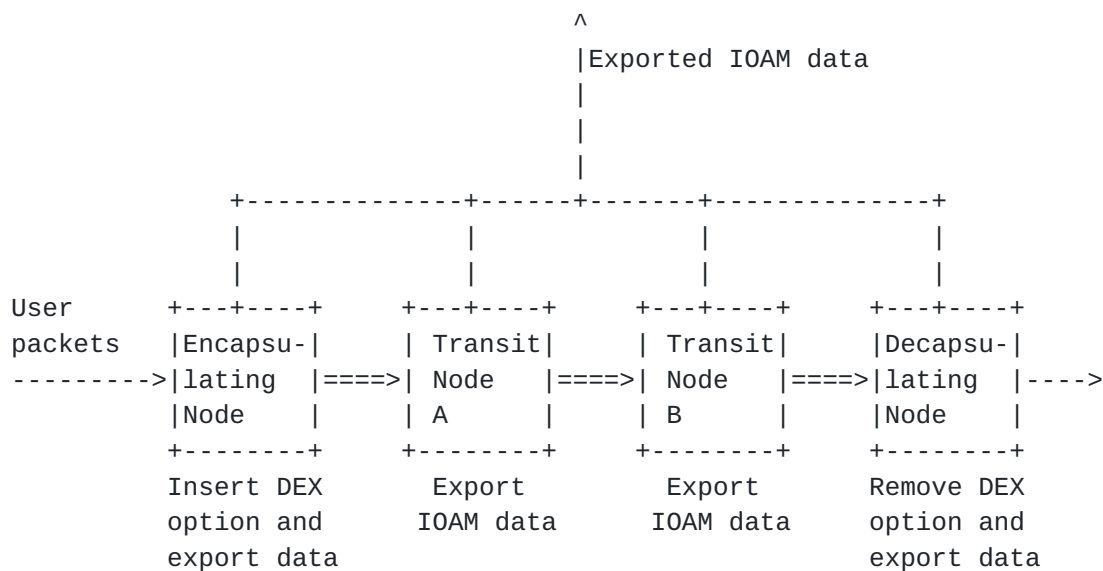
                   Figure 1: DEX Architecture

   The DEX option is used as a trigger to 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 the packet with the DEX option, and MAY export
      the requested IOAM data immediately.  The IOAM encapsulating node
      is the only type of node allowed to push the DEX option.

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

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

   As in [I-D.ietf-ippm-ioam-data], the DEX option may be incorporated
   into all or a subset of the traffic that is forwarded by the
   encapsulating node.  Moreover, IOAM nodes MAY export data for all
   traversing packets that carry the DEX option, or MAY selectively
   export data only for a subset of these packets.

   The DEX option specifies which data fields should be exported, as
   specified in Section 3.2.  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].

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
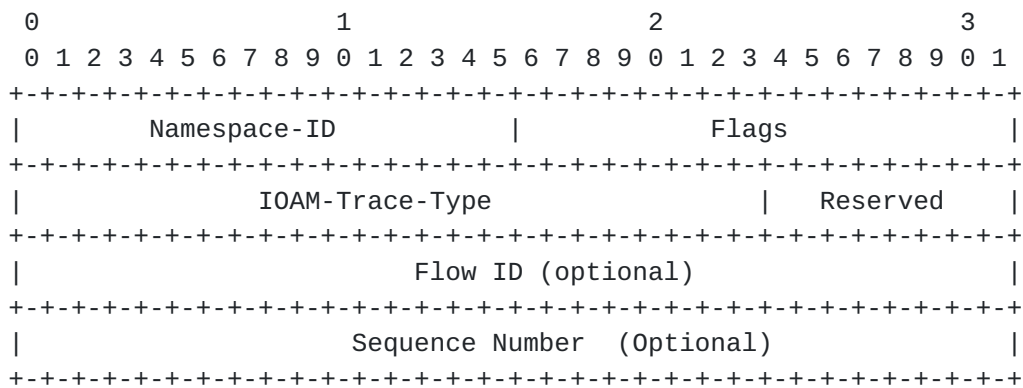present.

```
 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 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, rate limiting may be enabled so as to ensure that direct
   exporting is used at a rate that does not significantly affect the

   network bandwidth, and does not overload the receiving entity (or the
   source node in the case of loopback).  It should be possible to use
   each DEX on a subset of the data traffic, and to load balance the
   exported data among multiple receiving entities.

## 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, it should be
   possible for IOAM-enabled devices to limit the exported IOAM data to
   a configurable rate.

   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-11 (work in
              progress), November 2020.

   [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>.

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-03
              (work in progress), October 2020.

   [I-D.song-ippm-postcard-based-telemetry]
              Song, H., Zhou, T., Li, Z., Mirsky, G., Shin, J., and K.
              Lee, "Postcard-based On-Path Flow Data Telemetry using
              Packet Marking", draft-song-ippm-postcard-based-
              telemetry-08 (work in progress), October 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-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@huawei.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