Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: 9 January 2025

T. Graf
Swisscom
B. Claise
Huawei
A. Huang Feng
INSA-Lyon

8 July 2024

Export of On-Path Delay in IP Flow Information Export (IPFIX)

draft-ietf-opsawg-ipfix-on-path-telemetry-08

Abstract

This document ~~introduces~~ specifies new IP Flow Information Export (IPFIX) ~~i~~Information ~~elements~~ Elements to ~~expose~~ export the On-Path Telemetry measured delay on the In-situ OAM (IOAM) transit and decapsulation nodes.

Commenté [BMI1]: Might not be trivial to parse what's an "on-path delay" vs "path delay". More words are needed to have a meaningful title.

a mis en forme : Surlignage

Table of Contents

1. Introduction

Network operators ~~want~~ usually gather and maintain some forms of~~a~~ statistical delay view of their networks (or segments of their networks). That view is meant to help ~~They want to~~ understanding where in the network, for which customer ~~Traffic~~traffic or services, how much, and why abnormal delay is being ~~accummlated~~accumulated. ~~In order to answer why and where~~To that aim, delay-related data need~~s~~ to be reported ~~into~~ from devices covering both data ~~-~~and control~~-~~ planes ~~context~~. In order to understand which customer traffic is affected, delay-related data need~~s~~ to be reported into customer data-plane context. That enables network operators to quickly identify when the control-plane updates the current path with a different set of ~~next-hop~~ intermediate hops (that is, a change of the forwarding path)~~and therefore the forwarding path changes to different nodes~~ and interfaces, how the path delay changes for which customer traffic.

With On-Path Telemetry, described in the Network Telemetry Framework [RFC9232] and applied in In-situ OAM [I-D.ietf-ippm-ioam-deployment] and Alternate Marking Deployment Framework [I-D.ietf-ippm-alt-mark-deployment], the path delay between two endpoints ~~is~~ might be measured by inserting a timestamp in ~~the~~ packets.

At least two modes of On-Path Telemetry can be distinguished ~~between two modes~~. Passport mode, [RFC9197], where only the last hop in the forwarding path of the On-Path Telemetry domain exposes all the metrics, and postcard mode, [I-D.song-ippm-postcard-based-telemetry], where the metrics are also exposed in ~~the~~ transit nodes. In both modes the forwarding path exposes performance metrics allowing to determine how much delay has been accumulated on which hop.

This document defines four new IPFIX Information Elements (IEs), exposing the On-Path delay on IOAM transit and decapsulation nodes, following the postcard mode principles. Since these IPFIX IEs are performance metrics [RFC8911], they must be registered in the "IANA Performance Metric Registry [IANA-PERF-METRIC].

Following the guidelines for "Registered Performance Metric ~~requesters~~Requesters and R~~r~~eviewers" [RFC8911], the different characteristics of the performance metrics (Identifier, Name, URI, Status, Requester, Revision, Revision Date, Description, ~~etc~~etc.) must be clearly specified in the "IANA Performance Metric Registry [IANA-PERF-METRIC] in order for the measurement results ~~of measurements~~ using the Performance Metrics to be comparable even if they are performed ~~by~~ using different implementations and in different networks. These characteristics start by selecting a meaningful name, following the "MetricType_Method_SubTypeMethod_... Spec_Units_Output" naming convention (See Section 7.1.2 of

Commenté [BMI2]: As many delay metric flavors can be supported

Commenté [BMI3]: That is?

Commenté [BMI4]: In reference to what?

Commenté [BMI5]: Next-hop depends on which node we reason about.

I suspect the key point here is the change of intermediate nodes.

Commenté [BMI6]: Please update to RFC 9378 - In Situ Operations, Administration, and Maintenance (IOAM) Deployment (ietf.org)

Commenté [BMI7]: There is no such term in 9197.

Commenté [BMI8]: That spec expired. Please check if is needed here.

Commenté [BMI9]: I don't see in draft-song how "per-hop delay" is actually exposed. Please clarify where such details are detailed.

Commenté [BMI10]: The previous text does not adequately explain/motivate why IPFIX IEs are needed

Commenté [BMI11]: Cite base IPFIX RFCs

Commenté [BMI12]: Unless you have a better reference, I'm afraid I-D.song is normative then

Commenté [BMI13]: Redundant with the "must" in "they must be registered in the "IANA Performance Metric Registry [IANA-PERF-METRIC]."

Commenté [BMI14]: The results can be blindly comparable because this depends on these networks

Commenté [BMI15]: I don't parse this.

[RFC8911]).

Performance Metric	IPFIX Information Element
OWDelay_HybridType1_Passive_IP_RFC[RFC-to-be]_Seconds_Mean (TBD1)	PathDelayMeanDeltaMicroseconds (TBD5)
OWDelay_HybridType1_Passive_IP_RFC[RFC-to-be]_Seconds_Min (TBD2)	PathDelayMinDeltaMicroseconds (TBD6)
OWDelay_HybridType1_Passive_IP_RFC[RFC-to-be]_Seconds_Max (TBD3)	PathDelayMaxDeltaMicroseconds (TBD7)
OWDelay_HybridType1_Passive_IP_RFC[RFC-to-be]_Seconds_Sum (TBD4)	PathDelaySumDeltaMicroseconds (TBD8)

Commenté [BMI16]: Names of Information Elements MUST start with lowercase letters. (RFC 7012)

Commenté [BMI17]: Idem

Commenté [BMI18]: Idem

Commenté [BMI19]: Idem

Table 1: ~~Correspondance~~ Mapping ~~between~~ Between IPFIX IEs and Performance Metrics

The delay is measured by calculating the difference between the timestamp imposed with On-Path Telemetry in the packet at the IOAM encapsulation node and the timestamp exported in the IPFIX flow record from the IOAM transit and decapsulation nodes. The lowest, highest, mean, and/or the sum of measured path delay can be exported, thanks to the different IPFIX IEs~~ specifications~~.

Commenté [BMI20]: This assumes time sync is in place.

```
                  On-Path Telemetry Domain
           .........................................
           .                                       .
           .    D1                                 .
           . <------>                              .
           .                                       .
           .          D2                           .
           . <--------------------->               .
           .                                       .
           .                  D3                   .
           . <------------------------------------> .
           .                                       .
  (H1) ------ (R1) ------- (R2) ------- (R3) -------- (R4) ------ (H2)
  Host 1  Encapsulation   Transit      Transit    Decapsulation  Host 2
              Node         Node 1       Node 2        Node
           .                                       .
           .                                       .
           .........................................
```

Commenté [BMI21]: This might be trivial, but please indicate this about one-way delay.

Figure 1: Delay use case. Packets flow from host 1 to host 2.

~~On~~ In the ~~usecase~~ use case ~~showed~~ shown in Figure 1 using On-path Telemetry to export the delay metrics, the node R2 exports the delay D1, the node R3 exports the delay D2 and the decapsulation node R4 exports the total delay D3 using IPFIX.

Commenté [BMI22]: Some pointers to where the exact timestamping is described would be useful.

2. Terminology

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.

This document makes use of the terms defined in [RFC7011] and [I-D.ietf-ippm-ioam-deployment].

The following terms are used as defined in [RFC7011]:~~.~~

* IPFIX

* IPFIX Information Elements (IEs)

* Flow

* Flow Record

* Exporter

The following terms are used as defined in [RFC8911]:~~.~~

* Performance Metric

* Registered Performance Metric

* Performance Metrics Registry

The following terms are used as defined in Section 3 of [I-D.ietf-ippm-ioam-deployment]:~~.~~

* IOAM encapsulation node

* IOAM transit node

* IOAM decapsulation node

3. Performance Metrics

This section defines ~~and describes the~~ new performance metrics ~~by~~following ~~applying~~ the template defined in Section 11 of [RFC8911].

IANA Note (to be removed): RFC 8192 section 4 was taken as a guiding example.

3.1. IP One-Way Delay Hybrid Type I Passive Performance Metrics

This section specifies four performance metrics for the Hybrid Type I Passive assessment of IP One-Way Delay, to be registered in the "IANA Performance Metric Registry" [IANA-PERF-METRIC].

All column entries besides the Identifier~~ID~~, Name, Description, and Output Reference Method categories are the same; thus, this section defines four closely related performance metrics. As a result, IANA has

Commenté [BMI23]: To be udpated

Commenté [BMI24]: To be udpated

Commenté [BMI25]: To mirror the registry structure assigned corresponding ~~URLs~~ URIs to each of the four registered performance metrics.

3.1.1. Summary

This category includes multiple indexes of the registered performance metrics: the element ~~ID~~ Identifier and Metric Name.

3.1.1.1. ID (Identifier)

IANA ~~has~~ is requested to allocate~~d~~ the numeric Identifiers TBD1, TBD2, TBD~~TDB~~3, and TBD4
for the four Named Metric Entries in the following section.

RFC EDITOR NOTE: please replace TBD1, TBD2, ~~TDB3~~TBD3, and TBD4.

3.1.1.2. Name

TBD1: OWDelay_HybridType1_Passive_IP_RFC[RFC-to-be]_Seconds_Mean

TBD2: OWDelay_HybridType1_Passive_IP_RFC[RFC-to-be]_Seconds_Min

TBD3: OWDelay_HybridType1_Passive_IP_RFC[RFC-to-be]_Seconds_Max

TBD4: OWDelay_HybridType1_Passive_IP_RFC[RFC-to-be]_Seconds_Sum

~~RFC EDITOR NOTE: please replace TBD1, TBD2, TDB3, and TBD4.~~

3.1.1.3. URI

~~URL~~URI: https://www.iana.org/assignments/performance-metrics/
OWDelay_HybridType1_Passive_IP_RFC[RFC-to-be]_Seconds_Mean

~~URL~~URI: https://www.iana.org/assignments/performance-metrics/
OWDelay_HybridType1_Passive_IP_RFC[RFC-to-be]_Seconds_Min

~~URL~~URI: https://www.iana.org/assignments/performance-metrics/
OWDelay_HybridType1_Passive_IP_RFC[RFC-to-be]_Seconds_Max

~~URL~~URI: https://www.iana.org/assignments/performance-metrics/
OWDelay_HybridType1_Passive_IP_RFC[RFC-to-be]_Seconds_Sum

~~RFC EDITOR NOTE: please replace TBD1, TBD2, TDB3, and TBD4.~~

3.1.2. Description

* OWDelay_HybridType1_Passive_IP_RFC[RFC-to-be]_Seconds_Mean: This metric assesses the mean of one-way delays of all successfully forwarded IP packets constituting a single Flow ~~between two hosts~~. We consider the measurement of one-way delay based on a single Observation Point (OP) [RFC7011] somewhere in the network.

* OWDelay_HybridType1_Passive_IP_RFC[RFC-to-be]_Seconds_Min: This metric assesses the minimum of one-way delays of all successfully forwarded IP packets constituting a single Flow ~~between two hosts~~. We consider the measurement of one-way delay based on a single Observation Point (OP) [RFC7011] somewhere in the network.

Commenté [BMI26]: Where is this assigned? I don't see this in the registry. May be I'm not looking in the right place.

Commenté [BMI27]: Please check

Commenté [BMI28]: No need to repeat this each time.

Commenté [BMI29]: What is meant by "assess" here? This is not clear at least to me.

Commenté [BMI30]: That is covered by the definition of Flow

Commenté [BMI31]: Idem as previous comment

* OWDelay_HybridType1_Passive_IP_RFC[RFC-to-be]_Seconds_Max: This metric assesses the maximum of one-way delays of all successfully forwarded IP packets constituting a single Flow~~ between two hosts~~. We consider the measurement of one-way delay based on a single Observation Point (OP) [RFC7011] somewhere in the network.

* OWDelay_HybridType1_Passive_IP_RFC[RFC-to-be]_Seconds_Sum: This metric assesses the sum of one-way delays of all successfully forwarded IP packets constituting a single Flow~~ between two hosts~~. We consider the measurement of one-way delay based on a single Observation Point (OP) [RFC7011] somewhere in the network.

3.1.3. Change Controller

IETF

3.1.4. Version of Registry Format

1.0

3.2. Metric Definition

This category includes columns to prompt the entry of all necessary details related to the metric definition, including the immutable document reference and values of input factors, called "Fixed Parameters".

3.2.1. Reference Definition

Almes, G., Kalidindi, S., Zekauskas, M., and A. Morton, Ed., "A One-Way Delay Metric for IP Performance Metrics (IPPM)", STD 81, RFC 7679, DOI 10.17487/RFC7679, January 2016, <https://www.rfc-editor.org/info/rfc7679>. [RFC7679]

Morton, A. and E. Stephan, "Spatial Composition of Metrics", RFC 6049, DOI 10.17487/RFC6049, January 2011, <https://www.rfc-editor.org/info/rfc6049>. [RFC6049]

Section 3.4 of [RFC7679] provides the reference definition of the singleton (single value) one-way delay metric. Section 4.4 of [RFC7679] provides the reference definition expanded to cover a multi-value sample. Note that terms such as "singleton" and "sample" are defined in ~~section ~~Section 2 of [RFC2330].

With the OP [RFC7011] typically located between the hosts participating in the ~~IP ~~Flow, the one-way delay metric requires one individual measurement between the OP and sourcing host, such that the Spatial Composition [RFC6049] of the measurements yields a one-way delay singleton.

This document specifies how to export the performance metric ~~with~~using IPFIX.

a mis en forme : Surlignage

3.2.2. Fixed Parameters

None

3.3. Method of Measurement

This category includes columns for references to relevant sections of the RFC(s) and any supplemental information needed to ensure an unambiguous method for implementations.

3.3.1. Reference Methods

The foundational methodology for this metric is defined in ~~section~~ Section 4 of [RFC7323] using the Timestamps option with modifications that allow application at a mid-path OP [RFC7011].

3.3.2. Packet Stream Generation

The timestamp when the packet is being received at IOAM encapsulation node. Format depends on On-Path Telemetry implementation. For IOAM, Section 4.4.1 of [RFC9197] describes what kind of timestamps are supported. ~~Section~~ Sections 4.4.2.3 and 4.4.2.4 describe where the timestamp is being inserted. For the Enhanced Alternate Marking Method, Section 2 of [I-D.zhou-ippm-enhanced-alternate-marking] describes timestamp encoding and granularity.

Commenté [BMI32]: To be listed as normative. No?

3.3.3. Traffic Filtering (Observation) Details

Runtime Parameters (~~below~~in the following sections) may be used for Traffic Filtering.

3.3.4. Sampling Distribution

This metric requires a partial sample of all packets that qualify according to the Traffic Filter criteria.

3.3.5. Runtime Parameters and Data Format

Runtime Parameters are input factors that must be determined, configured into ~~the~~ a measurement system, and reported with the results for the context to be complete.

The hybrid type I metering parameters must be reported to provide the complete measurement context. As an example, if the IPFIX Metering Process is used, then the IPFIX Metering Process parameters (IPFIX Template Record, potential traffic filters, and potential sampling method and parameters) that generate the Flow Records must be reported to provide the complete measurement context. At a minimum, the following fields are required:

Src: The IP address of the host in the host A Role (format ipv4-address-no-zone value for IPv4 or ipv6-address-no-zone value for IPv6; see ~~section~~ Section 4 of [RFC6991]).

Dst: The IP address of the host in the host B Role (format ipv4-address-no-zone value for IPv4 or ipv6-address-no-zone value for IPv6; see ~~section~~ Section 4 of [RFC6991]).

T0: T time, the start of a measurement interval (format "date/time" as specified in Section 5.6 of [RFC3339]; see also "date-and-time" in Section 3 of [RFC6991]). The UTC Time Zone is required by Section 6.1 of [RFC2330]. When T0 is "all-zeros", a start time is unspecified and Tf is to be interpreted as the duration of the measurement interval. The start time is controlled through other means.

Tf: A time, the end of a measurement interval (format "date/time" as specified in Section 5.6 of [RFC3339]; see also "date-and-time" in Section 3 of [RFC6991])). The UTC Time Zone is required by Section 6.1 of [RFC2330]. When T0 is "all-zeros", an ending time and date is ignored and Tf is interpreted as the duration of the measurement interval.

3.3.6. Roles

host A: Launches ~~the~~ an IP packet to start the Flow. The Role of "host A" is synonymous with the IP address used at host A.

Commenté [BMI33]: ?

host B: Receives the IP packet to start the Flow. The Role of "host B" is synonymous with the IP address used at host B.

a mis en forme : Surlignage

Encapsulation Node: Receives the IP Flow packets and encapsulates the timestamp into the packet. The Role of "Encapsulation Node" is synonymous with the timestamp inserted in the packet.

a mis en forme : Surlignage

Transit Node: Receives the IP Flow packets and ~~measures~~ computes the delay between the timestamp in the packet and the timestamp when the packet was received.

Decapsulation Node: Receives the IP Flow packets and ~~measures~~ computes the delay between the timestamp in the packet and the timestamp when the packet was received and removes the IOAM header from the packet.

3.4. Output

This category specifies all details of the output of measurements using the metric.

3.4.1. Type

OWDelay Types are discussed in the subsections below.

3.4.2. Reference Definition

For all output types:

OWDelay_HybridType1_Passive_IP: The one-~~trip~~ way delay of one IP packet
is a Singleton

For each <statistic> Singleton one of the following subsections applies.

3.4.2.1. OWDelay_HybridType1_Passive_IP_RFC[RFC-to-be]_Seconds_Mean

Similar to Section 7.4.2.2 of [RFC8912], ~~The~~ the mean SHALL be calculated using the conditional distribution of all packets with a finite value of one-way delay (undefined delays are excluded) -- a single value, as follows:

Commenté [BMI34]: I would as similar reference for the other entries as the def is echoing what was defined there.

See ~~section~~ Section 4.1 of [RFC3393] for details on the conditional distribution to exclude undefined values of delay, and see section 5 of [RFC6703] for background on this analysis choice.

See ~~section~~ Section 4.2.2 of [RFC6049] for details on calculating this statistic; see also ~~section~~ Section 4.2.3 of [RFC6049].

Mean: The time value of the result is expressed in units of seconds, as a positive value of type decimal64 with fraction digits = 9 (see ~~section~~ Section 9.3 of [RFC6020]) with a resolution of 0.000000001 seconds (1.0 ns), and with lossless conversion to/from the 64-bit NTP timestamp as per ~~section~~ Section 6 of [RFC5905].

3.4.2.2. OWDelay_HybridType1_Passive_IP_RFC[RFC-to-be]_Seconds_Min

Similar to Section 7.4.2.3 of [RFC8912], the ~~The~~ minimum SHALL be calculated using the conditional distribution of all packets with a finite value of one-way delay (undefined delays are excluded) -- a single value, as follows:

Commenté [BMI35]: I would as similar reference for the other entries as the def is echoing what was defined there.

See S~~s~~ection 4.1 of [RFC3393] for details on the conditional distribution to exclude undefined values of delay, and see section 5 of [RFC6703] for background on this analysis choice.

See section 4.3.2 of [RFC6049] for details on calculating this statistic; see also S~~s~~ection 4.3.3 of [RFC6049].

Min: The time value of the result is expressed in units of seconds, as a positive value of type decimal64 with fraction digits = 9 (see section 9.3 of [RFC6020]) with a resolution of 0.000000001 seconds (1.0 ns), and with lossless conversion to/from the 64-bit NTP timestamp as per section 6 of [RFC5905].

3.4.2.3. OWDelay_HybridType1_Passive_IP_RFC[RFC-to-be]_Seconds_Max

Similar to Section 7.4.2.4 of [RFC8912], the~~The~~ maximum SHALL be calculated using the conditional distribution of all packets with a finite value of one-way delay (undefined delays are excluded) -- a single value, as follows:

Commenté [BMI36]: I would as similar reference for the other entries as the def is echoing what was defined there.

See section 4.1 of [RFC3393] for details on the conditional distribution to exclude undefined values of delay, and see section 5 of [RFC6703] for background on this analysis choice.

See section 4.3.2 of [RFC6049] for a closely related method for calculating this statistic; see also section 4.3.3 of [RFC6049]. The formula is as follows:

```
Max = (FiniteDelay[j])
such that for some index, j, where 1 <= j <= N
FiniteDelay[j] >= FiniteDelay[n] for all n
```

where all packets n = 1 through N have finite singleton delays.

Max: The time value of the result is expressed in units of seconds, as a positive value of type decimal64 with fraction digits = 9 (see section 9.3 of [RFC6020]) with a resolution of 0.000000001 seconds (1.0 ns), and with lossless conversion to/from the 64-bit NTP timestamp as per section 6 of [RFC5905].

3.4.2.4. OWDelay_HybridType1_Passive_IP_RFC[RFC-to-be]_Seconds_Sum

The sum SHALL be calculated using the conditional distribution of all packets with a finite value of one-way delay (undefined delays are excluded) -- a single value, as follows:

See S~~s~~ection 4.1 of [RFC3393] for details on the conditional distribution to exclude undefined values of delay, and see S~~s~~ection 5 of [RFC6703] for background on this analysis choice.

See S~~s~~ection 4.3.5 of [RFC6049] for details on calculating this statistic. However, in this case FiniteDelay or MaxDelay MAY be used.

Sum: The time value of the result is expressed in units of seconds, as a positive value of type decimal64 with fraction digits = 9 (see ~~section~~ Section 9.3 of [RFC6020]) with a resolution of 0.000000001 seconds (1.0 ns), and with lossless conversion to/from the 64-bit NTP timestamp as per S~~s~~ection 6 of [RFC5905].

Commenté [BMI37]: Why this ref is provided here?

3.4.2.5. Metric Units

* Mean

* Min

* Max

* Sum

The one-way delay of the IP Flow singleton is expressed in seconds.

3.4.2.6. Calibration

Passive Measurements at an OP could be calibrated against an Active Measurement at host A where the Active Measurement represents the ground truth.

3.4.3. Administrative Items

3.4.3.1. Status

Current

3.4.3.2. Requester

This RFC

RFC EDITOR NOTE: please replace This RFC text by the RFC issued from this document

3.4.3.3. Revision

1.0

3.4.3.4. Revision Date

RFC Date

3.4.4. Comments and Remarks

none

4. IPFIX Information Elements

This section ~~defines and describes~~specifies ~~the~~ the following new IPFIX IEs~~.~~:

PathDelayMeanDeltaMicroseconds
: 32-bit unsigned integer that identifies the mean path delay of all packets in the Flow, in microseconds, between the IOAM encapsulation ~~n ode~~node and the local node with the IOAM domain (either an IOAM transit node or an IOAM decapsulation node).

PathDelayMinDeltaMicroseconds
: 32-bit unsigned integer that identifies the lowest path delay of all packets in the Flow, in microseconds, between the IOAM encapsulation node and the local node with the IOAM domain (either an IOAM transit node or an IOAM decapsulation node).

PathDelayMaxDeltaMicroseconds
: 32-bit unsigned integer that identifies the highest path delay of all packets in the Flow, in microseconds, between the IOAM encapsulation node and the local node with the IOAM domain (either an IOAM transit node or an IOAM decapsulation node).

PathDelaySumDeltaMicroseconds
: 64-bit unsigned integer that identifies the sum of the path delay of all packets in the Flow, in microseconds, between the IOAM encapsulation node and the local node with the IOAM domain (either an IOAM transit node or an IOAM decapsulation node).

Commenté [BMI38]: This requires local caching and computation. Why not simply exporting observed timestamps/etc. and let the collector makes the computation?

Commenté [BMI39]: What is the measurement interval? Is it the same used for all exporting nodes?

5. Use Cases

The measured On-Path delay can be aggregated with Flow Aggregation as defined in [RFC7015] to the following device and control-plane dimensions to determine:

   * With node id and egressInterface(14)~~egressInterface(IE14)~~, on which node which
      logical egress interfaces have been contributing to how much
      delay.

   * With node id and egressPhysicalInterface(253), on which node which
      physical egress interfaces have been contributing to how much
      delay.

   * With ipNextHopIPv4Address(15) ~~ipNextHopIPv4Address(IE15)~~ or
ipNextHopIPv6Address(62)~~ipNextHopIPv6Address(IE62)~~, the
      forwarding path to which next-hop IP contributed to how much
      delay.

   * With mplsTopLabelIPv4Address(47)~~mplsTopLabelIPv4Address(IE47)~~ or
destinationIPv6Address(28) ~~destinationIPv6Address~~ and
      srhActiveSegmentIPv6 (495) from [RFC9487], the forwarding path to which
      MPLS top label IPv4 address or IPv6 destination address and SRv6
      active segment contributed to how much delay.

   * BGP communities [RFC1997] are often used for setting a path priority or
      service selection. With bgpDestinationExtendedCommunityList(488)
      or bgpDestinationCommunityList(485) or
      bgpDestinationLargeCommunityList(491) which group of prefixes
      accumulated at which node how much delay.

   * With destinationIPv4Address(13), destinationTransportPort(11),
      protocolIdentifier (4), and
sourceIPv4Address(8)~~sourceIPv4Address(IE8)~~, the forwarding
      path delay on each node from each IPv4 source address to a
      specific application in the network.

   ~~Taking~~ Let us consider the example depicted in ~~F~~figure 1 from ~~section~~
Section 1 as topology example. ~~Below example~~
   ~~T~~table 2 shows the aggregated delay per each node, ingressInterface,
   egressInterface, destinationIPv6Address and srhActiveSegmentIPv6.

ingress Interface	egress Interface	Node	destination IPv6Address	srhActive SegmentIPv6	Path Delay
271	276	R1	2001:db8::2	2001:db8::4	0 us
301	312	R2	2001:db8::3	2001:db8::4	22 us
22	27	R3	2001:db8::4	2001:db8::4	42 us
852	854	R4	2001:db8::4	2001:db8::4	122 us

Table 2: Example table of measured delay. Ascending by delay.

6. IANA Considerations

6.1. Performance Metrics

a mis en forme : Surlignage

Commenté [BMI40]: Why this one is cited? What about destinationIPv4Address?

Commenté [BMI41]: Please a consistent approach, either you cite the ref for all all IEs or you only refer to the registry.

For example [RFC8549] is not cited for communities.

Commenté [BMI42]: Cite a ref

Commenté [BMI43]: Why is this specific to v4?

Commenté [BMI44]: Indicate the IE elementID.

This document requests IANA to ~~create~~ add four new performance metrics under the "Performance Metrics" registry [RFC8911] with the four templates defined in ~~section~~ Section 3.

6.2. IPFIX Entities

This document requests IANA to ~~create~~ register new IPFIX IEs (see T~~t~~able 3) under the "IPFIX Information Elements" registry [RFC7012] available at "IANA Performance Metric Registry [IANA-PERF-METRIC] and assign the following initial code points.

Element ID	Name
TBD5	PathDelayMeanDeltaMicroseconds
TBD6	PathDelayMinDeltaMicroseconds
TBD7	PathDelayMaxDeltaMicroseconds
TBD8	PathDelaySumDeltaMicroseconds

Table 3: ~~Creates~~ New IPFIX IEs in the "IPFIX Information Elements" R~~r~~egistry

~~Note to the RFC-Editor:~~

~~* Please replace TBD5 - TBD8 with the values allocated by IANA~~

~~* Please replace all instances of [RFC-to-be] in this section with the RFC number assigned to this document~~

6.2.1. ~~PathDelayMeanDeltaMicroseconds~~pathDelayMeanDeltaMicroseconds

Name: ~~PathDelayMeanDeltaMicroseconds~~pathDelayMeanDeltaMicroseconds

ElementID: TBD5

Description: This Information Element identifies the mean path delay of all packets in the Flow, in microseconds, between the IOAM encapsulation node and the local node with the IOAM domain (either an IOAM transit node or an IOAM decapsulation node), according to OWDelay_HybridType1_Passive_IP_RFC[RFC-to-be]_Seconds_Mean in the IANA Performance Metric Registry

Abstract Data Type: unsigned32

Data Type Semantics: deltaCounter

Reference: [RFC-to-be], OWDelay_HybridType1_Passive_IP_RFC[RFC-to-be]_Seconds_Mean in the IANA Performance Metric Registry.

Commenté [BMI45]: I guess you meant the IPFIX registry group. Right?

Commenté [BMI46]: This is too specific. The IE can be applicable in other contexts

Commenté [BMI47]: This belongs more to the Additional Info

6.2.2. ~~PathDelayMinDeltaMicroseconds~~pathDelayMinDeltaMicroseconds

Name: ~~PathDelayMinDeltaMicroseconds~~pathDelayMinDeltaMicroseconds

ElementID: TBD6

Description: This Information Element identifies the lowest path delay of all packets in the Flow, in microseconds, between the IOAM encapsulation node and the local node with the IOAM domain (either an IOAM transit node or an IOAM decapsulation node), according to the OWDelay_HybridType1_Passive_IP_RFC[RFC-to-be]_Seconds_Min in the IANA Performance Metric Registry.

Abstract Data Type: unsigned32

Data Type Semantics: deltaCounter

Reference: [RFC-to-be], OWDelay_HybridType1_Passive_IP_RFC[RFC-to-be]_Seconds_Min in the IANA Performance Metric Registry.

6.2.3. ~~PathDelayMaxDeltaMicroseconds~~pathDelayMaxDeltaMicroseconds

Name: ~~PathDelayMaxDeltaMicroseconds~~pathDelayMaxDeltaMicroseconds

ElementID: TBD7

Description: This Information Element identifies the highest path delay of all packets in the Flow, in microseconds, between the IOAM encapsulation node and the local node with the IOAM domain (either an IOAM transit node or an IOAM decapsulation node), according to OWDelay_HybridType1_Passive_IP_RFC[RFC-to-be]_Seconds_Max in the IANA Performance Metric Registry.

Abstract Data Type: unsigned32

Data Type Semantics: deltaCounter

Reference: [RFC-to-be], OWDelay_HybridType1_Passive_IP_RFC[RFC-to-be]_Seconds_Max in the IANA Performance Metric Registry.

6.2.4. p~~P~~athDelaySumDeltaMicroseconds

Name: ~~PathDelaySumDeltaMicroseconds~~pathDelaySumDeltaMicroseconds

ElementID: TBD8

Description: This Information Element identifies the sum of the path delay of all packets in the Flow, in microseconds, between the IOAM encapsulation node and the local node with the IOAM domain (either an IOAM transit node or an IOAM decapsulation node), according to OWDelay_HybridType1_Passive_IP_RFC[RFC-to-be]_Seconds_Sum in the IANA Performance Metric Registry.

Abstract Data Type: unsigned64

Data Type Semantics: deltaCounter

Commenté [BMI48]: Idem as previous IE

Commenté [BMI49]: Idem as previous IE

a mis en forme : Surlignage

a mis en forme : Surlignage

Reference: [RFC-to-be], OWDelay_HybridType1_Passive_IP_RFC[RFC-to-be]_Seconds_Sum in the IANA Performance Metric Registry.

7. Operational Considerations

7.1. Time Accuracy

The same recommendation as defined in S~~s~~ection 4.5 of [RFC5153] for IPFIX applies in terms of clock precision to this document as well.

7.2. Mean Delay

The mean (average) path delay can be calculated by dividing the ~~PathDelaySumDeltaMicroseconds~~pathDelaySumDeltaMicroseconds(TBD5) by the packetDeltaCount(2) at the IPFIX data collection in order to offload the IPFIX Exporter from calculating the mean for every Flow at export time.

7.3. Reduced-size encoding

Unsigned64 has been chosen as type for ~~PathDelaySumDeltaMicroseconds~~pathDelaySumDeltaMicroseconds to support cases with large delay numbers and where many packets are being accounted. As an example, a specific ~~flow~~ Flow ~~record~~ Record with path delay of 100 microseconds ~~can not~~cannot observe more than 42949 packets without overflowing the unsigned32 counter. The procedure described in Section 6.2 of [RFC7011] ~~could be~~ may be applied to reduce network bandwidth between the IPFIX Exporter and Collector if unsigned32 would be large enough without wrapping around.

7.4. IOAM Application

This document is applicable in IOAM to the Edge-to-Edge and Direct Exporting Option-Type.

In case of Edge-to-Edge Option-Type, as described in Section 4.6 of [RFC9197], by setting bits 2 and 3, timestamps can be encoded as defined in ~~Section~~ Sections 4.4.2.3 and 4.4.2.4 of [RFC9197].

In case of Direct Exporting Option-Type, as described in Section 2 of [I-D.ahuang-ippm-dex-timestamp-ext], by setting Extension-Flags 2 and 3, timestamps can be encoded as defined in ~~Section~~ Sections 4.4.2.3 and 4.4.2.4 of [RFC9197].

For the Enhanced Alternate Marking Method, Section 2 of [I-D.zhou-ippm-enhanced-alternate-marking] defines that within the metaInfo a nano second timestamp can be encoded in the encapsulation node and be read at the intermediate and decapsulation node to calculate the on-path delay. [RFC9343] defines how this can be ~~appied~~applied to the IPv6 data-plane and [I-D.fz-spring-srv6-alt-mark] defines how this can be ~~appied~~applied to the Segment Routing Header in SRv6.

8. Security Considerations

There are no significant extra security considerations regarding the allocation of these new IPFIX IEs compared to [RFC7012].

Commenté [BMI50]: The export of these IEs is critical as it may trigger service path ajustement that may themselves lead to service distorsion

This is unique to these IEs. I would elaborate on the guards against blindly trusting these.

9. Implementation Status

Note to the RFC-Editor: Please remove this section before publishing.

9.1. FD.io VPP

INSA Lyon implemented the following IEs as part of a prototype in the FD.io VPP (Vector Packet Processing) platform:

* PathDelayMeanDeltaMicroseconds

* PathDelayMaxDeltaMicroseconds

* PathDelayMinDeltaMicroseconds

* PathDelaySumDeltaMicroseconds

The open source code can be obtained here: [INSA-Lyon-VPP] and was validated at the IETF 116 hackathon.

9.2. Huawei VRP

Huawei implemented the following IEs as part of a a production implementation in the VRP platform:

* PathDelayMeanDeltaMicroseconds

* PathDelayMaxDeltaMicroseconds

* PathDelayMinDeltaMicroseconds

* PathDelaySumDeltaMicroseconds

The implementation was validated at the IETF 116 hackathon.

9.3. Fluvia

NTT Com implemented the following IEs in the Fluvia Exporter:

* PathDelayMeanDeltaMicroseconds

* PathDelayMaxDeltaMicroseconds

* PathDelayMinDeltaMicroseconds

* PathDelaySumDeltaMicroseconds

The open source code can be obtained here: [NTT-Fluvia] and was validated at the IETF 118 hackathon.

9.4. Pmacct Data Collection

Paolo Lucente implemented the IE PathDelayMeanDeltaMicroseconds by dividing IE PathDelaySumDeltaMicroseconds by IE packetDeltaCount in the open source Network Telemetry data collection project pmacct.

The source code can be obtained here: [Paolo-Lucente-Pmacct] and was validated at the IETF 116 hackathon.

10. Acknowledgements

The authors would like to thank Al Morton, Greg Mirsky and Giuseppe Fioccola for their review and valuable comments.

11. References

11.1. Normative References

[RFC7012] Claise, B., Ed. and B. Trammell, Ed., "Information Model for IP Flow Information Export (IPFIX)", RFC 7012, DOI 10.17487/RFC7012, September 2013, <https://www.rfc-editor.org/info/rfc7012>.

[RFC8911] Bagnulo, M., Claise, B., Eardley, P., Morton, A., and A. Akhter, "Registry for Performance Metrics", RFC 8911, DOI 10.17487/RFC8911, November 2021, <https://www.rfc-editor.org/info/rfc8911>.

11.2. Informative References

Commenté [BMI51]: Please double check your list. Some entries are definitely normative (2119, for example).

I have doubt about postcard and so on

[I-D.ahuang-ippm-dex-timestamp-ext] Feng, A. H., Francois, P., Claise, B., and T. Graf, "Timestamp extension for In Situ Operations, Administration, and Maintenance (IOAM) Direct Export", Work in Progress, Internet-Draft, draft-ahuang-ippm-dex-timestamp-ext-00, 15 February 2023, <https://datatracker.ietf.org/doc/html/draft-ahuang-ippm-dex-timestamp-ext-00>.

[I-D.fz-spring-srv6-alt-mark] Fioccola, G., Zhou, T., Cociglio, M., Mishra, G. S., wang, X., and G. Zhang, "Application of the Alternate Marking Method to the Segment Routing Header", Work in Progress, Internet-Draft, draft-fz-spring-srv6-alt-mark-08, 8 February 2024, <https://datatracker.ietf.org/doc/html/draft-fz-spring-srv6-alt-mark-08>.

[I-D.ietf-ippm-alt-mark-deployment] Fioccola, G., Keyi, Z., Graf, T., Nilo, M., and L. Zhang, "Alternate Marking Deployment Framework", Work in Progress, Internet-Draft, draft-ietf-ippm-alt-mark-deployment-01, 3 July 2024, <https://datatracker.ietf.org/doc/html/draft-ietf-ippm-alt-mark-deployment-01>.

[I-D.ietf-ippm-ioam-deployment] Brockners, F., Bhandari, S., Bernier, D., and T. Mizrahi, "In-situ OAM Deployment", Work in Progress, Internet-Draft, draft-ietf-ippm-ioam-deployment-05, 4 January 2023, <https://datatracker.ietf.org/doc/html/draft-ietf-ippm-ioam-deployment-05>.

[I-D.song-ippm-postcard-based-telemetry]
Song, H., Mirsky, G., Zhou, T., Li, Z., Graf, T., Mishra, G. S., Shin, J., and K. Lee, "On-Path Telemetry using Packet Marking to Trigger Dedicated OAM Packets", Work in Progress, Internet-Draft, draft-song-ippm-postcard-based-telemetry-16, 2 June 2023, <https://datatracker.ietf.org/doc/html/draft-song-ippm-postcard-based-telemetry-16>.

[I-D.zhou-ippm-enhanced-alternate-marking]
Zhou, T., Fioccola, G., Liu, Y., Cociglio, M., Pang, R., Xiong, L., Lee, S., and W. Li, "Enhanced Alternate Marking Method", Work in Progress, Internet-Draft, draft-zhou-ippm-enhanced-alternate-marking-15, 27 May 2024, <https://datatracker.ietf.org/doc/html/draft-zhou-ippm-enhanced-alternate-marking-15>.

[IANA-PERF-METRIC]
"IANA Performance Metric Registry", <https://www.iana.org/assignments/performance-metrics/performance-metrics.xhtml>.

[INSA-Lyon-VPP]
"INSA Lyon, FD.io VPP implementation", <https://github.com/network-analytics/vpp-srh-onpath-telemetry>.

[NTT-Fluvia]
"NTT Com, Fluvia Exporter", <https://github.com/nttcom/fluvia/>.

[Paolo-Lucente-Pmacct]
"Paolo Lucente, Pmacct open source Network Telemetry Data Collection", <https://github.com/pmacct/pmacct>.

[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>.

[RFC2330] Paxson, V., Almes, G., Mahdavi, J., and M. Mathis, "Framework for IP Performance Metrics", RFC 2330, DOI 10.17487/RFC2330, May 1998, <https://www.rfc-editor.org/info/rfc2330>.

[RFC3339] Klyne, G. and C. Newman, "Date and Time on the Internet: Timestamps", RFC 3339, DOI 10.17487/RFC3339, July 2002, <https://www.rfc-editor.org/info/rfc3339>.

[RFC3393] Demichelis, C. and P. Chimento, "IP Packet Delay Variation Metric for IP Performance Metrics (IPPM)", RFC 3393, DOI 10.17487/RFC3393, November 2002, <https://www.rfc-editor.org/info/rfc3393>.

[RFC5153] Boschi, E., Mark, L., Quittek, J., Stiemerling, M., and P. Aitken, "IP Flow Information Export (IPFIX) Implementation Guidelines", RFC 5153, DOI 10.17487/RFC5153, April 2008, <https://www.rfc-editor.org/info/rfc5153>.

[RFC5905] Mills, D., Martin, J., Ed., Burbank, J., and W. Kasch, "Network Time Protocol Version 4: Protocol and Algorithms Specification", RFC 5905, DOI 10.17487/RFC5905, June 2010, <https://www.rfc-editor.org/info/rfc5905>.

[RFC6020] Bjorklund, M., Ed., "YANG - A Data Modeling Language for the Network Configuration Protocol (NETCONF)", RFC 6020, DOI 10.17487/RFC6020, October 2010, <https://www.rfc-editor.org/info/rfc6020>.

[RFC6049] Morton, A. and E. Stephan, "Spatial Composition of Metrics", RFC 6049, DOI 10.17487/RFC6049, January 2011, <https://www.rfc-editor.org/info/rfc6049>.

[RFC6703] Morton, A., Ramachandran, G., and G. Maguluri, "Reporting IP Network Performance Metrics: Different Points of View", RFC 6703, DOI 10.17487/RFC6703, August 2012, <https://www.rfc-editor.org/info/rfc6703>.

[RFC6991] Schoenwaelder, J., Ed., "Common YANG Data Types", RFC 6991, DOI 10.17487/RFC6991, July 2013, <https://www.rfc-editor.org/info/rfc6991>.

[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>.

[RFC7015] Trammell, B., Wagner, A., and B. Claise, "Flow Aggregation for the IP Flow Information Export (IPFIX) Protocol", RFC 7015, DOI 10.17487/RFC7015, September 2013, <https://www.rfc-editor.org/info/rfc7015>.

[RFC7323] Borman, D., Braden, B., Jacobson, V., and R. Scheffenegger, Ed., "TCP Extensions for High Performance", RFC 7323, DOI 10.17487/RFC7323, September 2014, <https://www.rfc-editor.org/info/rfc7323>.

[RFC7679] Almes, G., Kalidindi, S., Zekauskas, M., and A. Morton, Ed., "A One-Way Delay Metric for IP Performance Metrics (IPPM)", STD 81, RFC 7679, DOI 10.17487/RFC7679, January 2016, <https://www.rfc-editor.org/info/rfc7679>.

[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>.

[RFC9197] Brockners, F., Ed., Bhandari, S., Ed., and T. Mizrahi, Ed., "Data Fields for In Situ Operations, Administration, and Maintenance (IOAM)", RFC 9197, DOI 10.17487/RFC9197, May 2022, <https://www.rfc-editor.org/info/rfc9197>.

[RFC9232] Song, H., Qin, F., Martinez-Julia, P., Ciavaglia, L., and A. Wang, "Network Telemetry Framework", RFC 9232, DOI 10.17487/RFC9232, May 2022,

<https://www.rfc-editor.org/info/rfc9232>.

[RFC9343] Fioccola, G., Zhou, T., Cociglio, M., Qin, F., and R. Pang, "IPv6 Application of the Alternate-Marking Method", RFC 9343, DOI 10.17487/RFC9343, December 2022, <https://www.rfc-editor.org/info/rfc9343>.

[RFC9487] Graf, T., Claise, B., and P. Francois, "Export of Segment Routing over IPv6 Information in IP Flow Information Export (IPFIX)", RFC 9487, DOI 10.17487/RFC9487, November 2023, <https://www.rfc-editor.org/info/rfc9487>.