IPPM Working Group
Internet-Draft
Intended status: Standards Track
Expires: 31 August 2026

X. He
China Telecom
F. Brockners
Cisco
H. Song
Futurewei
G. Fioccola
Huawei
A. Wang
China Telecom

27 February 2026

IOAM Direct Exporting (DEX) Option Extensions for Incorporating the Alternate-Marking Method
draft-he-ippm-ioam-dex-extensions-incorporating-am-04

Abstract

In situ Operations, Administration, and Maintenance (IOAM) is used for recording and collecting operational and telemetry information. Specifically, passport-based IOAM allows telemetry data generated by each node along the path to be pushed into data packets when they traverse the network, while postcard-based IOAM allows IOAM data generated by each node to be directly exported without being pushed into in-flight data packets. The Alternate-Marking method is used to measure performance metrics on live traffic, such as packet loss, delay, and jitter. This document extends IOAM Direct Export (DEX) Option-Type to integrate the Alternate-Marking Method into IOAM to augment IOAM in performance measurement.

Table of Contents

1. Introduction

IOAM [RFC9197], which defines two possible IOAM Trace Option-Types: Pre-allocated Trace, Incremental Trace, is used for monitoring traffic in the network and for incorporating IOAM data fields into in-flight data packets. IOAM [RFC9197] is known as the passport mode, in which each node on the path can add telemetry data to the user packets (i.e., stamps the passport). IOAM Direct Export (DEX) [RFC9326] is used as a trigger for IOAM nodes to directly export IOAM data to a receiving entity such as a collector, analyzer, or

controller. IOAM DEX is also referred as the postcard mode, in which each node directly exports the telemetry data using an independent packet (i.e., sends a postcard) while the user packets are unmodified.

The limitation of the passport mode lies in that if a packet is lost on the path, the collected IOAM data will also be lost. So the passport mode such as IOAM Trace Option-Type is unable to monitor packet loss and its location.

IOAM DEX Option-Type can complement IOAM Trace Option-Type. Even if a packet is lost on the path, the partial data collected is still available. By correlating the data from different nodes, the number of the lost packets can be counted accurately and packet drop location can also be pinpointed.

The Alternate-Marking [RFC9341] technique has been proven to work well to perform packet loss, delay, and jitter measurements on live traffic. RFC9343 describes how the Alternate-Marking Method can be used to measure performance metrics in IPv6. It defines an Extension Header Option to encode Alternate-Marking information in both the Hop-by-Hop Options Header and Destination Options Header.

This document presents the problems and challenges currently faced by IOAM in measuring performance metrics such as packet loss, delay, and jitter. In order to augment IOAM in performance measurement, IOAM DEX Option-Type is extended to incorporate the Alternate-Marking method into IOAM.

2. Conventions

2.1. Requirements 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:

DEX: Direct Exporting

FlowMonID: Flow Monitoring Identification

IOAM: In situ Operation, Administration, and Maintenance

MPN: Measurement Period Number

OAM: Operation, Administration, and Maintenance

SN: Sequence Number

3. Problem Statement

Although IOAM DEX Option-Type can complement IOAM Trace Option-Type for monitoring packet loss, some issues have to be considered as follows.

* Measurement accuracy: In theory, when an IOAM encapsulating node incorporates the DEX Option-Type into all the traffic it forwards, the fidelity of packet loss measurement can be guaranteed. If only a very small traffic subset or an excessively low traffic sampling rate is implemented on an encapsulating node, loss measurement results will not reflect the actual packet loss, since packet loss caused by instantaneous congestion such as microburst may not be detected. Therefore, a higher traffic sampling rate can help improve measurement accuracy.

* Performance impact: If an IOAM encapsulating node incorporates the DEX Option-Type into all the traffic of interest it forwards, it may lead to an excessive amount of exported data, which may significantly occupy network bandwidth and overload the receiving entity. On the other hand, because the IOAM data of the same user packet is generated by every node along the path, the receiving entity needs more processing overhead to correlate these data for packet loss calculation. The more user packets being measured, the more processing overhead is required. Therefore, an IOAM encapsulating node that supports the DEX Option-Type must also support the ability to incorporate the DEX Option-Type selectively into a subset of the packets that are forwarded by the IOAM encapsulating node.

* Measurement methodology: When using the Alternate-Marking method, traffic flows are split into consecutive blocks: each block represents a measurable entity unambiguously recognizable by all network devices along the path. In contrast, according to IOAM DEX Option-Type, every IOAM node directly exports an IOAM data to a receiving entity when every user packet is forwarded, and the collected IOAM data are not split into independent measurement blocks. It is the receiving entity’s responsibility to determine the measurement period for performance metrics such as packet loss, delay, and jitter, which is not beneficial to a unified measurement methodology.

4. Integrate the Alternate-Marking Method into IOAM DEX Option

To address the issues and challenges mentioned in Section 3, IOAM needs to be augmented to implement performance measurement. The Alternate-Marking method has been widely employed in operator networks. By integrating the Alternate-Marking method into IOAM DEX Option-Type, the benefits gained include:

* Guaranteed measurement accuracy: When implementing performance measurement using the Alternate-Marking method, an IOAM encapsulating node may incorporate the DEX Option-Type into all packets of the measured user traffic it forwards, and could color all the traffic of interest, not a subset of the packets, thus the fidelity of performance measurement such as packet loss can be guaranteed.

* Minimal performance impact: When using the Alternate-Marking method for loss measurement, in Hop-by-Hop mode, every node along the path exports only a packet carrying counter value of each measurement block including a batch of packets; In End-to-End mode, only the IOAM encapsulating node and the IOAM decapsulating node export a packet carrying counter value of each measurement block. As a result, it significantly mitigates the network and the receiving entity. Furthermore, compared to IOAM DEX Option-Type, the receiving entity needs much less processing overhead to correlate these counter values for packet loss computation.

* Unified measurement methodology: When using the Alternate-Marking method, traffic flows are split into consecutive blocks: each block represents a measurable entity unambiguously recognizable by all network devices along the path, thus the measurement period is completely determined by the encapsulating node. The receiving entity does not need to concern about how to determine measurement period, but only compute the result for each measurement period. It is beneficial to a unified measurement methodology.

* Unified packet header format: Furthermore, by incorporating the Alternate-Marking method into IOAM DEX Option-Type, only unique packet header encapsulation format is used for both IOAM trace monitoring and performance measurement, thus simplifing the complexity of forwarding chips.

5. The Extended DEX Option-Type Format

The format of the Extended DEX Option-Type is depicted in Figure 1. All fields are same as DEX Option-Type format defined in RFC9326 except the Reserved field. The Extended DEX Option-Type format uses the most significant 2 bits of the Reserved field.

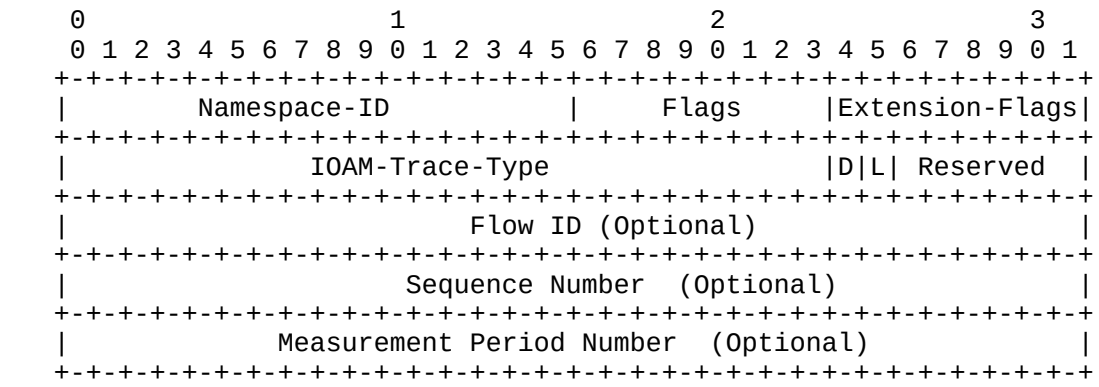

Figure 1: The Extended DEX Option-Type Format

Where the fields are defined as follows:

Namespace-ID: 16-bit identifier of the IOAM namespace, as defined in [RFC9197].

Flags: 8-bit field, comprised of 8 1-bit subfields,as defined in [RFC9326]. Flags are allocated by IANA.

Extension-Flags: 8-bit field, comprised of 8 1-bit subfields, as defined in [RFC9326]. Extension-Flags are allocated by IANA. Every bit in the Extension-Flag field that is set to 1 indicates the existence of a corresponding optional 4-octet field. Bit 0 (the most significant bit) and bit 1 in the registry are allocated by [RFC9326], which are specified as Flow ID and Sequence Number of the monitored traffic, respectively. Bit 4 (TBA) is specified as Measurement Period Number in this document. An IOAM node that receives an extended DEX Option-Type with an unknown flag set to 1 MUST ignore the corresponding optional field.

IOAM-Trace-Type: 24-bit identifier that specifies which IOAM data types are used and the corresponding IOAM-Data-Fields should be exported. The format of this field is as defined in [RFC9197].

L: 1-bit Loss flag for Packet Loss Measurement as described in Section 6.1.

D: 1-bit Delay flag for Single Packet Delay Measurement as described in Section 6.2.

Reserved: 6-bit field, reserved for future use. These bits MUST be set to zero on transmission and ignored on receipt.

Optional fields: The optional fields, if present, reside after the Reserved field. The order of the optional fields is according to the order of the respective bits, starting from the most significant bit, that are enabled in the Extension-Flags field. Each optional field is 4 octets long.

Flow ID: An optional 32-bit field representing the 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 and exported to the receiving entity by the forwarding nodes. The Flow ID can be used to correlate the exported data of the same flow from multiple nodes and from multiple packets. Flow ID values are expected to be allocated in a way that avoids collisions. For example, random assignment of Flow ID values can be subject to collisions, while centralized allocation can avoid this problem. The collision probability (CP) for random allocation of Flow ID as well as the Flow ID allocation solution in the distributed way is detailed in Setion 7.

Sequence Number: An optional 32-bit sequence number, starting from 0 and incremented by 1 for each packet from the same flow at the encapsulating node that includes the Extended DEX option. The Sequence Number, when combined with the Flow ID, provides a convenient approach to correlate the exported data from the same user packet.

Measurement Period Number(MPN): An optional 32-bit field representing the measurement period number of the monitored flow, starting from 0 and incremented by 1 for the specified flow with the same Flow ID. The field is set at the encapsulating node and exported to the receiving entity by the forwarding nodes. The MPN, when combined with the Flow ID, provides a convenient approach to correlate the exported data of the same flow during the same measurement period from multiple nodes.

6. The IOAM Operation

The Extended DEX Option-Type SHOULD support the three IOAM operation modes: only IOAM trace monitoring; only performance measurement; hybrid.

* Only IOAM trace monitoring: As DEX Option-Type does, an IOAM encapsulating node that supports the Extended DEX Option-Type MUST support the ability to incorporate the Extended DEX Option-Type selectively into a subset of the monitored packets that are forwarded by the IOAM encapsulating node. At the same time, it MUST set the corresponding bit flag to 1 in IOAM-Trace-Type field of the Extended DEX Option-Type so that each node along the path

needs to generate the specified IOAM data exported to the receiving entity. Also, it MUST set the L flag and D flag of the Extended DEX Option-Type to zero on transmit and ignored by the monitoring nodes.

* Only performance measurement: To ensure the fidelity of performance measurement, an IOAM encapsulating node MUST incorporate the Extended DEX Option-Type into all packets of the measured user traffic it forwards. Like the Alternate-Marking method [RFC9343], for packet loss measurement, it MUST switch the value of the L bit between 0 and 1 after a fixed number of packets or according to a fixed timer; for packet delay measurement, Single-Marking or Double-Marking methodology can be adopted by switching the value of the L bit or D bit between 0 and 1. But it MUST set all 24 bits flag to 0 in IOAM-Trace-Type field of the Extended DEX Option-Type so that each node along the path does not need to generate the IOAM data exported to the receiving entity.

* Hybrid: To perform both IOAM trace monitoring and performance measurement concurrently, an IOAM encapsulating node MUST incorporate the Extended DEX Option-Type into all the traffic of interest it forwards. For performance measurement, an IOAM encapsulating node MUST mark each packet it forwards in L flag and D flag of the Extended DEX Option-Type; for IOAM trace monitoring, only a subset of the packets are selected by an IOAM encapsulating node. For every selected packet, an IOAM encapsulating node MUST set corresponding bit flag to 1 in IOAM-Trace-Type field of the Extended DEX Option-Type so that each node along the path needs to generate the specified IOAM data exported to the receiving entity; for all the other packets not selected, an IOAM encapsulating node MUST set all 24 bits flag to 0 in IOAM- Trace-Type field of the Extended DEX Option-Type, such that each node along the path does not need to generate the IOAM data exported to the receiving entity.

6.1. Packet Loss Measurement

The measurement of the packet loss is detailed in [RFC9341]and [RFC9343]. The packets of the flow identified by Flow ID are grouped into batches, and all the packets within a batch are marked by setting the L bit (Loss flag) to a same value. The source node (IOAM encapsulating node) can switch the value of the L bit between 0 and 1 after a fixed number of packets or according to a fixed timer, and this depends on the implementation. The source node is the only one that marks the packets to create the batches, while the intermediate nodes only read the marking values and identify the packet batches. By counting the number of packets in each batch and comparing the values measured by different network nodes along the path, it is

possible to measure the packet loss that occurred in any single batch between any two nodes. Each batch represents a measurable entity recognizable by all network nodes along the path, which export the counter value of this batch along with the Flow ID and the MPN (if present) to the receiving entity (e.g., the collector).

By employing the optional MPN field, it provides a simpler and accurate approach for the receiving entity to correlate the packet counter values belonging to the same measurement block (or measurement period) from different nodes, especially in the case where milliseconds-level measurement block for real-time network performance monitoring is required, e.g., Precision Availability Metrics (PAM) described in RFC9544.

6.2. Packet Delay Measurement

Delay metrics may be calculated using the following two possibilities:

Single-Marking Methodology: This approach uses only the L bit to calculate both packet loss and delay. In this case, the D flag MUST be set to zero on transmit and ignored by the monitoring nodes. The alternation of the values of the L bit can be used as a time reference to calculate the delay. Whenever the L bit changes and a new batch starts, a network node can store the timestamp of the first packet of the new batch; that timestamp can be compared with the timestamp of the first packet of the same batch on a second node to compute packet delay. However, this measurement is accurate only when the first packet of the new batch has not experienced packet loss or packet reordering.

Double-Marking Methodology: This approach is to use the first marking with the L bit to create the alternate batch and, within the batches identified by the L bit, a second marking with the D bit set to 1 is used to select the packets for measuring delay. The D bit creates a new set of marked packets that are fully identified over the network so that a forwarding node can store and export the timestamps of these packets, and these timestamps can be compared with the timestamps of the same packets on a second node to compute packet delay values for each packet. Compared to selecting a single double-marked packet for each batch, which may lead to an invalid delay result if this double-marked packet is lost, selecting multiple double-marked packets for each measurement batch provides a robust and accurate method for packet delay measurement. Furthermore, the average delay for multiple double-marked packets can also be obtained for every measurement period.

Sequence Number can be used to identify multiple timestamps in different packets that pertain to the same measurement block in case of packet reordering. Also, it can be used to identify which double-marked packet is lost.

In summary, the approach with Double-Marking is better than the approach with Single-Marking. In the implementation, the timestamps along with Flow ID, MPN and Sequence Number (if present) can be sent out to the receiving entity that is responsible for the calculation.

7. Flow Identification

The Flow Identification (Flow ID) identifies the flow to be measured and is required for some general reasons, which is described in Section 5.3 of [RFC9343]. [RFC9343] uses 20-bit FlowMonID to determine a monitored flow within the measurement domain. Compared to the FlowMonID, the Flow ID in this document is a 32-bit field, which amplifies the FlowMonID space by 4096 times. Accordingly, a chance of collision is greatly reduced in a distributed way.

When the 32-bit Flow ID is used for every source node, if there are N edge nodes (source nodes) in a large-scale operator network, and each source node can generate a unique Flow ID for every measured flow independently and randomly in a distributed way. Assuming that each node randomly generates M different Flow IDs from the available K flow identification space, then the total possible sample space (PSS) is

the Nth power of C (K, M) (1)

and the total possible sample space without overlapping (PSSno)is

C1 (K, M)*C2 (K-M, M)*....*CN (k-(N-1)M, M) (2)

Theoritically, the collision probability (CP) is calculated as:

CP=1-PSSno/PSS (3)

Take K=32nd power of 2 as an example, which corresponds to 32-bit Flow ID space. When the number N and M are given different values, we can obtain the corresponding CP values shown in the following table.

	N=100	N=100	N=200	N=200	N=200	N=200
32-bit Flow ID	M=100	M=200	M=200	M=300	M=400	M=500
CP	0.0115	0.0453	0.1692	0.3410	0.5235	0.6860

It is not difficult to observe that as the number of concurrent monitored flows increases, the collision probability is rapidly increasing. As shown in the table, when generating 10000 concurrent flows, the CP is 0.0115; when generating 100000 concurrent flows, the CP rises to 0.6860. If K=20th power of 2 is taken, which corresponds to 20-bit Flow ID space, when generating 10000 concurrent flows, we can calculate the collision probability will drastically rises to approximately 100%. In practical deployment scenarios of large-scale networks, the concurrent measurement flows could reach orders of magnitude of 100000 or even higher, thus the collision probability will rise sharply.

It is preferred that Flow ID be assigned by the centralized controller. Since the controller knows the network topology, it can allocate the value properly to guarantee the uniqueness of Flow ID allocation.

In some cases where the centralized controller is not available and the distributed way must be adopted, every source node (encapsulating node) needs to allocate Flow ID independently. In order to avoid the collision, Flow ID field may be devided into two sub-fields: NodeID and FlowMonID. NodeID is assigned uniquely in measurement domain (by the unified planning) and FlowMonID is assigned randomly and uniquely in a device. The length allocation of the two sub-fields depends on practical implementation, for example, NodeID uses 20 bits and FlowMonID uses 12 bits, or both use an average of 16 bits.

8. IANA Considerations

8.1. IOAM Type

The "IOAM Option-Type" registry is defined in Section 7.1 of [RFC9197].

IANA is requested to allocate the following code point from the "IOAM Option-Type" registry as follows:

Code Point: TBA.

Name: IOAM Extended DEX Option Type.

Description: Direct exporting.

Reference: This document.

If possible, IANA is requested to allocate code point 5(TBA-type).

8.2. IOAM DEX Extension-Flags

IANA has created the "IOAM DEX Extension-Flags" registry. This registry includes 8 flag bits. Bit 0 (the most significant bit) and bit 1 in the registry are allocated by [RFC9326]. Bit 2 and Bit 3 in the registry are allocated by [RFC9630].

IANA is requested to allocate the following bit from the "IOAM DEX Extension-Flags" registry as follows:

Bit: 4 (TBA).

Description: Measurement Period Number (MPN).

Reference: This document.

9. Performance Considerations

The Extended DEX Option-Type triggers IOAM data (including IOAM trace data and performance measurement data) to be collected and/or to be exported to a receiving entity. In some cases, this may impact the receiving entity’s performance.

Therefore, the performance impact of these exported packets is limited by taking two measures: at the encapsulating nodes by selective DEX encapsulation and at the transit nodes by limiting exporting rate, which are detailed in [RFC9326]. 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.

When performance measurement is implemented based on the Alternate-Marking Method, in Hop-by-Hop mode for loss measurement, every node along the path exports only a packet carrying counter value of each measurement block including a batch of packets; In End-to-End mode for loss measurement, only the IOAM encapsulating node and the IOAM decapsulating node export a packet carrying counter value of each measurement block. Meanwhile, an IOAM encapsulating node only needs to select a very small subset of the packets that are forwarded for IOAM trace monitoring (e.g., 1/10000 of all the traffic), so the amount of exported data is significantly reduced to mitigate the network and the receiving entity. Compared with IOAM DEX Option-Type

for packet loss calculation, due to a significant reduction in the number of exported packets, the receiving entity needs much less processing overhead to correlate these counter values for packet loss computation.

10. Security Considerations

The security considerations of IOAM in general are discussed in [RFC9197], the security considerations of IOAM DEX Option-Type are discussed in [RFC9326], and the security considerations of the Alternate-Marking method are discussed in [RFC9394]. There are not additional security considerations in this Extended IOAM DEX Option-Type.

11. References

11.1. Normative References

[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>.

[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>.

[RFC9326] Song, H., Gafni, B., Brockners, F., Bhandari, S., and T. Mizrahi, "In Situ Operations, Administration, and Maintenance (IOAM) Direct Exporting", RFC 9326, DOI 10.17487/RFC9326, November 2022, <https://www.rfc-editor.org/info/rfc9326>.

[RFC9341] Fioccola, G., Ed., Cociglio, M., Mirsky, G., Mizrahi, T., and T. Zhou, "Alternate-Marking Method", RFC 9341, DOI 10.17487/RFC9341, December 2022, <https://www.rfc-editor.org/info/rfc9341>.

[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>.

11.2. Informative References

[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>.

[RFC9486] Bhandari, S., Ed. and F. Brockners, Ed., "IPv6 Options for In Situ Operations, Administration, and Maintenance (IOAM)", RFC 9486, DOI 10.17487/RFC9486, September 2023, <https://www.rfc-editor.org/info/rfc9486>.

Authors' Addresses

Xiaoming He
China Telecom
Email: hexm4@chinatelecom.cn

Frank Brockners
Cisco
Email: fbrockne@cisco.com

Haoyu Song
Futurewei
Email: haoyu.song@futurewei.com

Giuseppe Fioccola
Huawei
Email: giuseppe.fioccola@huawei.com

Aijun Wang
China Telecom
Email: wangaj3@chinatelecom.cn