INTERNET-DRAFT N. Elkins
Inside Products
G. Fioccola
Telecom Italia
M. Ackermann
BCBS Michigan
R. Hamilton
Intended Status: Proposed Standard Chemical Abstract Services
Expires: April 24, 2018 October 21, 2018

IPv6 Marking and Performance and Diagnostic Metrics (MPDM)
draft-fear-ippm-mpdm-02

Abstract

To assess performance problems, this document describes optional headers embedded in each packet that provide marking, sequence numbers and timing information as a basis for measurements. Such measurements may be interpreted in real-time or after the fact. This document specifies the IPv6 Marking and Performance and Diagnostic Metrics (M-PDM) Hop-byHop and Destination Options extension headers.

Table of Contents

1 Background

To assess performance problems, measurements based on marking, sequence numbers and timing may be embedded in each packet. Such measurements may be interpreted in real-time or after the fact.

As defined in RFC8200 [RFC8200], destination options are carried by the IPv6 Destination Options extension header. Destination options include information that need be examined only by the IPv6 node given as the destination address in the IPv6 header, not by routers or "middle boxes".

RFC8200 [RFC8200] additionally defines the IPv6 Hop-by-Hop (HBH) Options extension header. This header may be processed and examined by end nodes, routers and "middle boxes".

This document specifies both the Marking Performance and Diagnostic Metrics (M-PDM) destination option as well as the M-PDM HBH Option.

1.1 Terminology

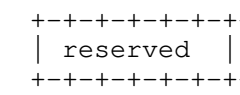
The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

1.2 Rationale for defined solution

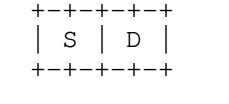
The M-PDM Destination Option and M-PDM Hop-By-Hop Option are a combining of PDM [RFC8250] with Marking [RFC8321] to obtain path and middle box information.

The Marking Field is designed as:

The 2 currently used bits from the 8 bit Marking field are designated as Mark Field (MF).



Mark Field (MF) is:



1.2.1 Alternate Marking Method Operation

[RFC8321] describes in detail the methodology, that we briefly illustrate also here.

1.2.1.1 Single Mark Measurement

As explained in the [RFC8321], marking can be applied to delineate blocks of packets based either on equal number of packets in a block or based on equal time interval. The latter method offers better control as it allows better account for capabilities of downstream nodes to report statistics related to batches of packets and, at the same time, time resolution that affects defect detection interval.

If the Single Mark measurement used, then the D flag MUST be set to zero on transmit and ignored by monitoring point.

The S flag is used to create alternate flows to measure the packet loss by switching value of the S flag. Delay metrics MAY be calculated with the alternate flow using any of the following methods:

- First/Last Batch Packet Delay calculation: timestamps are collected based on order of arrival so this method is sensitive to packet loss and re-ordering.

- Average Packet Delay calculation: an average delay is calculated by considering the average arrival time of the packets within a single block. This method only provides single metric for the duration of the block and it doesn't give information about the delay distribution.

1.2.2.2 Double Mark Measurement

Double Mark method allows more detailed measurement of delays for the monitored flow but it requires more nodal and network resources. If the Double Mark method used, then the S flag MUST be used to create the alternate flow. The D flag MUST be used to mark single packets to measure delay jitter.

The first marking (S flag alternation) is needed for packet loss and also for average delay measurement. The second marking (D flag is put to one) creates a new set of marked packets that are fully identified and dedicated for delay. This method is useful to have not only the average delay but also to know more about the statistic distribution of delay values.

1.3 IPv6 Transition Technologies

In the path to full implementation of IPv6, transition technologies such as translation or tunneling may be employed. It is possible that an IPv6 packet containing M-PDM may be dropped if using IPv6 transition technologies. For example, an implementation using a translation technique (IPv6 to IPv4) which does not support or recognize the IPv6 Destination Options extension header or a new HBH option may simply drop the packet rather than translating it without the extension header.

It is also possible that some devices in the network may not correctly handle multiple IPv6 Extension Headers, including the IPv6 Destination Option. For example, adding the PDM header to a packet may push the layer 4 information to a point in the packet where it is not visible to filtering logic, and may be dropped. This kind of situation is expected to become rare over time.

2 Measurement Information Derived from PDM

Each packet contains information about the sender and receiver. In IP protocol, the identifying information is called a "5-tuple".

The 5-tuple consists of:

```
SADDR : IP address of the sender
SPORT : Port for sender
DADDR : IP address of the destination
DPORT : Port for destination
PROTC : Protocol for upper layer (ex. TCP, UDP, ICMP, etc.)
```

The PDM contains the following base fields:

```
PSNTP    : Packet Sequence Number This Packet
PSNLR    : Packet Sequence Number Last Received
DELTATLR : Delta Time Last Received
DELTATLS : Delta Time Last Sent
```

This information, combined with the 5-tuple, allows the measurement of the following metrics:

1. Round-trip delay
2. Server delay

These are further described in RFC8250 [RFC8250].

Performance measurements described in [RFC8321] are allowed.

3 Marking and Performance and Diagnostic Metrics (M-PDM) Destination Option Layout

3.1 Destination Options Header

The IPv6 Destination Options Header is used to carry information that needs to be examined only by a packet's destination node(s). The Destination Options Header is identified by a Next Header value of 60 in the immediately preceding header and is defined in RFC8200 [RFC8200]. The IPv6 Marking and Performance and Diagnostic Metrics Destination Option (M-PDM) is implemented as an IPv6 Option carried in the Destination Options Header. M-PDM does not require time synchronization.

3.2 Marking and Performance and Diagnostic Metrics (M-PDM) Destination Option

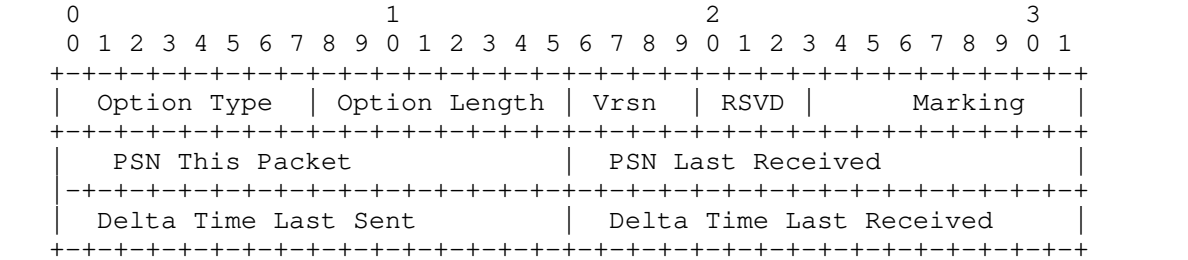
3.2.1 M-PDM Layout

The IPv6 Marking and Performance and Diagnostic Metrics Destination Option (M-PDM) contains the following fields:

```
PSNTP    : Packet Sequence Number This Packet
PSNLR    : Packet Sequence Number Last Received
DELTATLR : Delta Time Last Received
DELTATLS : Delta Time Last Sent
```

PDM has alignment requirements. Following the convention in IPv6, these options are aligned in a packet so that multi-octet values within the Option Data field of each option fall on natural boundaries (i.e., fields of width n octets are placed at an integer multiple of n octets from the start of the header, for n = 1, 2, 4, or 8) [RFC8200].

The M-PDM destination option is encoded in type-length-value (TLV) format as follows:



Option Type

TBD = 0xXX (TBD) [To be assigned by IANA] [RFC2780]

In keeping with RFC8200 [RFC8200], the two high-order bits of the Option Type field are encoded to indicate specific processing of the option; for the PDM destination option, these two bits MUST be set to 00.

The third high-order bit of the Option Type field specifies whether or not the Option Data of that option can change en route to the packet's final destination.

In M-PDM, the value of the third high-order bit MUST be 0.

Option Length

8-bit unsigned integer. Length of the option, in octets, excluding the Option Type and Option Length fields. This field MUST be set to 10.

Version

4-bit unsigned integer.

Reserved

4-bit unsigned integer.

Marking

8-bit unsigned integer. (2 currently used - 6 reserved)

Packet Sequence Number This Packet (PSNTP)

16-bit unsigned integer. This field will wrap. It is intended for use while analyzing packet traces.

This field is initialized at a random number and incremented monotonically for each packet of the session flow of the IP stack. The random-number initialization is intended to make it harder to spoof and insert such packets.

Packet Sequence Number Last Received (PSNLR)

16-bit unsigned integer. This is the PSNTP of the packet last received by the IP stack.

This field is initialized to 0.

Delta Time Last Sent (DELTATLS)

16-bit unsigned integer.

Delta Time Last Sent = (receive time packet n - send time packet (n - 1))

Delta Time Last Received (DELTATLR)

16-bit unsigned integer.

Delta Time Last Received = (send time packet n - receive time packet (n - 1))

3.2.2 Base Unit for Time Measurement

Fixed base. TBD. [More information needs to be added here.]

3.3 Header Placement

The M-PDM Destination Option is placed as defined in RFC8200 [RFC8200]. There may be a choice of where to place the Destination Options header. If using ESP mode, please see section 3.4 of this document for placement of the M-PDM Destination Options header.

For each IPv6 packet header, the M-PDM MUST NOT appear more than once. However, an encapsulated packet MAY contain a separate M-PDM associated with each encapsulated IPv6 header.

3.4 Header Placement Using IPSec ESP Mode

IPSec Encapsulating Security Payload (ESP) is defined in [RFC4303] and is widely used. Section 3.1.1 of [RFC4303] discusses placement of Destination Options Headers.

The placement of M-PDM is different depending on if ESP is used in tunnel or transport mode.

In ESP case, no 5-tuple is available, as there are no port numbers. ESP flow should be identified only by using SADDR, DADDR and PROTOC. The SPI numbers SHOULD be ignored when considering the flow over which M-PDM information is measured.

3.4.1 Using ESP Transport Mode

Note that Destination Options may be placed before or after ESP or both. If using M-PDM in ESP transport mode, M-PDM MUST be placed after the ESP header so as not to leak information.

3.4.2 Using ESP Tunnel Mode

Note that Destination Options may be placed before or after ESP or both in both the outer set of IP headers and the inner set of IP headers. A tunnel endpoint that creates a new packet may decide to use M-PDM independent of the use of M-PDM of the original packet to enable delay measurements between the two tunnel endpoints.

3.5 Implementation Considerations

3.5.1 M-PDM Activation

An implementation should provide an interface to enable or disable the use of M-PDM. This specification recommends having M-PDM off by default.

M-PDM MUST NOT be turned on merely if a packet is received with an M-PDM header. The received packet could be spoofed by another device.

3.5.2 M-PDM Timestamps

The M-PDM timestamps are intended to isolate wire time from server or host time, but may necessarily attribute some host processing time to network latency.

RFC2330 [RFC2330] "Framework for IP Performance Metrics" describes two notions of wire time in section 10.2. These notions are only defined in terms of an Internet host H observing an Internet link L at a particular location:

+ For a given IP packet P, the 'wire arrival time' of P at H on L is the first time T at which any bit of P has appeared at H's observational position on L.

+ For a given IP packet P, the 'wire exit time' of P at H on L is the first time T at which all the bits of P have appeared at H's observational position on L.

This specification does not define the exact H's observing position on L. That is left for the deployment setups to define. However, the position where PDM timestamps are taken SHOULD be as close to the physical network interface as possible. Not all implementations will be able to achieve the ideal level of measurement.

3.6 Dynamic Configuration Options

If the M-PDM destination options extension header is used, then it MAY be turned on for all packets flowing through the host, applied to an upper-layer protocol (TCP, UDP, SCTP, etc), a local port, or IP address only. These are at the discretion of the implementation.

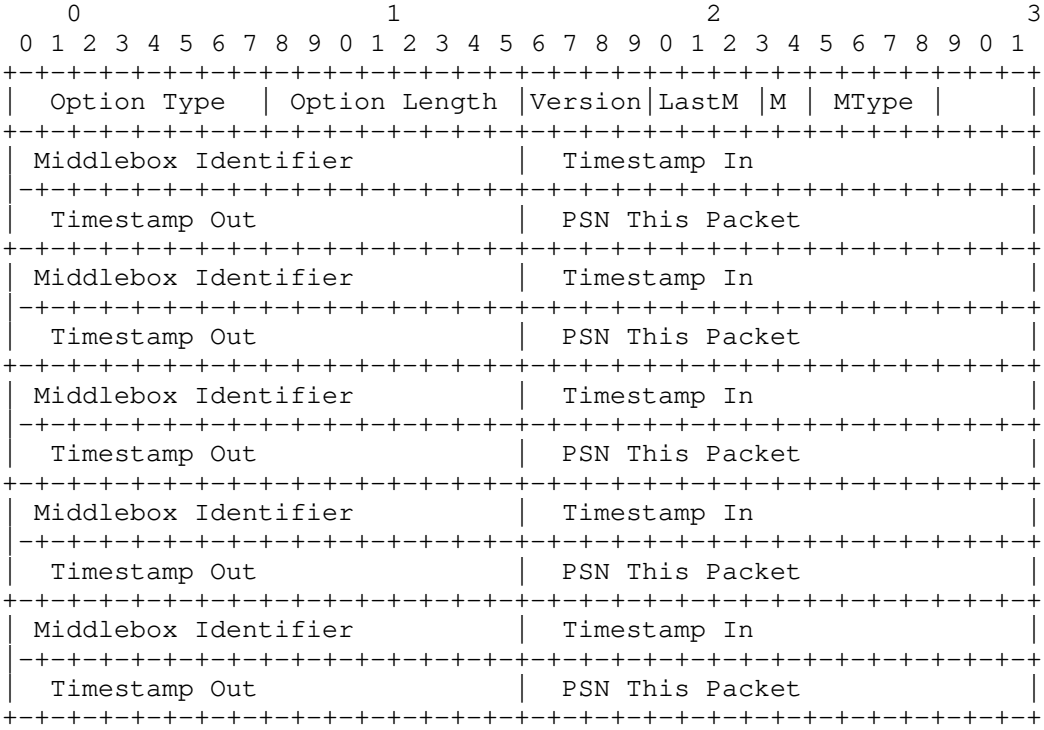
3.7 Information Access and Storage

Measurement information provided by M-PDM may be made accessible for higher layers or the user itself. Similar to activating the use of M-PDM, the implementation may also provide an interface to indicate if received.

M-PDM information may be stored, if desired. If a packet with M-PDM information is received and the information should be stored, the upper layers may be notified. Furthermore, the implementation should define a configurable maximum lifetime after which the information can be removed as well as a configurable maximum amount of memory that should be allocated for PDM information.

4 M-PDM HBH Option

The M-PDM Hop-by-Hop option is encoded in type-length-value (TLV) format. It has an alignment requirement of 4n + 2. (See [IPv6, Section 4.2] for discussion of option alignment.) The option has the following format:



Option Type

TBD = 0xXX (TBD) [To be assigned by IANA] [RFC2780]

In keeping with RFC 8200 [RFC8200], the two high-order bits of the Option Type field are encoded to indicate specific processing of the option; for the M-PDM HBH option, these two bits MUST be set to 00.

The third high-order bit of the Option Type field specifies whether or not the Option Data of that option can change en route to the packet's final destination.

In M-PDM, the value of the third high-order bit MUST be 0.

Option Length

8-bit unsigned integer. Length of the option, in octets, excluding the Option Type and Option Length fields. This field MUST be set to 10.

Version

4-bit unsigned integer.

Last Middlebox

4-bit unsigned integer. Indicates which middlebox number was last done. For example, 3 would indicate that this is the third middlebox. This field could be used to quickly find which set of data to fill. If there have been more than 5 middleboxes, then wrapping will happen and fields will get overwritten.

Marking

2-bit unsigned integer.

Marking Type (M-Type)

4-bit unsigned integer. This indicates the type of marking method being used for the timestamp.

If marking is not used, then the timestamp will be when the packet left the IP interface on this middlebox.

If marking method is used, then this field will contain:

1 - the timestamp of the first packet of a marked batch 2 - the average timestamp of the packets of a batch 3 - a double-marked packet

RSVD

2-bit unsigned integer

Middle Box Identifier

16-bit unsigned integer.

This field MUST be zero if not used. The zeros are intended to make it harder to leak data via the HBH header.

This could be some portion of the IPv4 or IPv6 address or the router ID. [Note to readers: any suggestions for this field are most welcome!]

Timestamp In

16-bit unsigned integer. This can be the timestamp of the packet received by the IP interface on this middlebox. If marking method is used, it can identify the timestamp of the first packet of a marked batch or the average timestamp of the packets of a batch or a double-marked packet, depending on which method is used to perform delay measurements.

This field is initialized to 0.

This timestamp is the delta in nanoseconds from the initial starting timestamp of January 1, 2019 00:00:00.0000000000.

See the section on HBH Timestamps for more on this measurement.

Timestamp Out

16-bit unsigned integer. This can be the timestamp of the packet left the IP interface on this middlebox. If marking method is used, it can identify the timestamp of the first packet of a marked batch or the average timestamp of the packets of a batch or a double-marked packet, depending on which method is used to perform delay measurements.

This field is initialized to 0.

This timestamp is the delta in nanoseconds from the initial starting timestamp of January 1, 2019 00:00:00.0000000000.

See the section on HBH Timestamps for more on this measurement.

Packet Sequence Number This Packet (PSNTP)

16-bit unsigned integer. This field will wrap. It is intended for use while analyzing packet traces.

This field is initialized at a random number and incremented monotonically for each packet of the session flow of the IP stack.

The random-number initialization is intended to make it harder to spoof and insert such packets.

4.1 HBH Timestamps In and Out

The timestamp fields will contain the 16 high-order or most significant bits of the delta between a fixed starting value of January 1, 2019 00:00:00.0000000000 and the current time at the middlebox.

For more on truncation of timestamp values, please see [TCPM].

5 Security Considerations

M-PDM may introduce some new security weaknesses.

5.1 Resource Consumption and Resource Consumption Attacks

M-PDM needs to calculate the deltas for time and keep track of the sequence numbers. This means that control blocks which reside in memory may be kept at the end hosts per 5-tuple.

A limit on how much memory is being used SHOULD be implemented. Without a memory limit, any time a control block is kept in memory, an attacker can try to misuse the control blocks to cause excessive resource consumption. This could be used to compromise the end host.

M-PDM as a Destination is used at the end hosts and memory is used only at the end host M-PDM as an HBH header is used at routers or middle boxes.

5.2 Pervasive monitoring

Since M-PDM passes in the clear, a concern arises as to whether the data can be used to fingerprint the system or somehow obtain information about the contents of the payload.

Let us discuss fingerprinting of the end host first. It is possible that seeing the pattern of deltas or the absolute values could give some information as to the speed of the end host - that is, if it is a very fast system or an older, slow device. This may be useful to the attacker. However, if the attacker has access to PDM, the attacker also has access to the entire packet and could make such a deduction based merely on the time frames elapsed between packets WITHOUT PDM.

As far as deducing the content of the payload, in terms of the

application level information such as web page, user name, user password and so on, it appears to us that PDM is quite unhelpful in this regard. Having said that, the ability to separate wire-time from processing time may potentially provide an attacker with additional information. It is conceivable that an attacker could attempt to deduce the type of application in use by noting the server time and payload length. Some encryption algorithms attempt to obfuscate the packet length to avoid just such vulnerabilities. In the future, encryption algorithms may wish to obfuscate the server time as well.

5.3 M-PDM as a Covert Channel

PDM provides a set of fields in the packet which could be used to leak data. But, there is no real reason to suspect that PDM would be chosen rather than another part of the payload or another Extension Header.

A firewall or another device could sanity check the fields within the PDM but randomly assigned sequence numbers and delta times might be expected to vary widely. The biggest problem though is how an attacker would get access to PDM in the first place to leak data. The attacker would have to either compromise the end host or have Man in the Middle (MitM). It is possible that either one could change the fields. But, then the other end host would get sequence numbers and deltas that don't make any sense.

It is conceivable that someone could compromise an end host and make it start sending packets with PDM without the knowledge of the host. But, again, the bigger problem is the compromise of the end host. Once that is done, the attacker probably has better ways to leak data.

Having said that, if a PDM aware middle box or an implementation (destination host) detects some number of "nonsensical" sequence numbers or timing information, it could take action to block, discard, or alert on this traffic.

5.4 Timing Attacks

The fact that PDM can help in the separation of node processing time from network latency brings value to performance monitoring. Yet, it is this very characteristic of PDM which may be misused to make certain new type of timing attacks against protocols and implementations possible.

Depending on the nature of the cryptographic protocol used, it may be possible to leak the credentials of the device. For example, if an

attacker can see that PDM is being used, then the attacker might use PDM to launch a timing attack against the keying material used by the cryptographic protocol.

An implementation may want to be sure that PDM is enabled only for certain ip addresses, or only for some ports. Additionally, the implementation SHOULD require an explicit restart of monitoring after a certain time period (for example for 1 hour), to make sure that PDM is not accidentally left on after debugging has been done etc.

Even so, if using PDM, a user "Consent to be Measured" SHOULD be a pre-requisite for using PDM. Consent is common in enterprises and with some subscription services. The actual content of "Consent to be Measured" will differ by site but it SHOULD make clear that the traffic is being measured for quality of service and to assist in diagnostics as well as to make clear that there may be potential risks of certain vulnerabilities if the traffic is captured during a diagnostic session.

6 IANA Considerations

This draft requests an Destination Option Type assignment with the act bits set to 00 and the chg bit set to 0 from the Destination Options and Hop-by-Hop Options sub-registry of Internet Protocol Version 6 (IPv6) Parameters [ref to RFCs and URL below.

http://www.iana.org/assignments/ipv6-parameters/ipv6-parameters.xhtml#ipv6-parameters-2

Hex Value	Binary Value act chg rest	Description	Reference
TBD	TBD	Performance and Diagnostic Metrics (M-PDM)	[This draft]

7 References

7.1 Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC2780] Bradner, S. and V. Paxson, "IANA Allocation Guidelines For Values In the Internet Protocol and Related Headers", BCP 37, RFC 2780, March 2000.

[RFC4303] Kent, S, "IP Encapsulating Security Payload (ESP)", RFC 4303, December 2005.

[RFC8200] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", STD 86, RFC 8200, DOI 10.17487/RFC8200, July 2017, <https://www.rfc-editor.org/info/rfc8200>.

[RFC8250] Elkins, N., Ackermann, M. and Hamilton, R. "IPv6 Performance and Diagnostic Metrics (PDM) Destination Option", RFC 8250, September 2017.

7.2 Informative References

[RFC2330] Paxson, V., Almes, G., Mahdavi, J., and M. Mathis, "Framework for IP Performance Metrics", RFC 2330, May 1998.

[RFC8321] Fioccola, G. et al, "Alternate-Marking Method for Passive and Hybrid Performance Monitoring", RFC 8321, January 2018.

[TCPM] Scheffenegger, R., Kuehlewind, M., and B. Trammell, "Encoding of Time Intervals for the TCP Timestamp Option", Work in Progress, draft-trammell-tcpm-timestamp-interval-01, July 2013

Acknowledgments

The authors would like to C.M. Heard for his review.

Authors' Addresses

Nalini Elkins
Inside Products, Inc.
36A Upper Circle
Carmel Valley, CA 93924
United States
Phone: +1 831 659 8360
Email: nalini.elkins@insidethestack.com
http://www.insidethestack.com

Giuseppe Fioccola
Telecom Italia
Via Reiss Romoli, 274
Torino 10148
Italy
Email: giuseppe.fioccola@telecomitalia.it

Michael S. Ackermann
Blue Cross Blue Shield of Michigan
P.O. Box 2888
Detroit, Michigan 48231
United States
Phone: +1 310 460 4080
Email: mackermann@bcbsm.com

Robert M. Hamilton
Chemical Abstracts Service
A Division of the American Chemical Society
2540 Olentangy River Road
Columbus, Ohio 43202
United States of America
Phone: +1 614 447 3600 x2517
Email: rhamilton@cas.org