INTERNET-DRAFT

N. Elkins
Inside Products
R. Hamilton
Chemical Abstracts Service
M. Ackermann
BCBS Michigan

Intended Status: Proposed Standard
Expires: April 2015
October 10, 2014

IPv6 Performance and Diagnostic Metrics Destination Option
draft-elkins-6man-ipv6-pdm-dest-option-08

Abstract

To measure performance and to diagnose performance and connectivity problems, metrics embedded in each packet are critical for timely and accurate problem resolution. Such diagnostics may be interpreted in real-time or after the fact. The base metrics are: packet sequence number and packet timing. Metrics derived from these will be described separately. This document describes a new implementation of the existing IPv6 Destination Options extension header, the Performance and Diagnostic Metrics (PDM) Destination Options extension header.

Copyright and License Notice

Table of Contents

1 Introduction

To measure performance and to diagnose performance and connectivity problems, metrics embedded in each packet are critical for timely and accurate problem resolution. Such diagnostics may be interpreted in real-time or after the fact. The base metrics are: packet sequence number and packet timing. Metrics derived from these will be described separately. This document describes a new implementation of an existing IPv6 Destination Options extension header, the Performance and Diagnostic Metrics (PDM) Destination Options extension header.

For the rationale and usage of the PDM extension header, see "IPPM Considerations for the IPv6 PDM Destination Option" [ELK-IPPM]. [ELK-IPPM] discusses measurement of end-user Quality of Experience (QoE) as well as the details of the scaling factors used. The current document describes the fields in the PDM header itself.

As defined in RFC2460 [RFC2460], destination options are carried by the IPv6 Destination Options extension header. Destination options include optional information that need be examined only by the IPv6 node given as the destination address in the IPv6 header, not by routers or other "middle boxes". This document specifies a new destination option, the Performance and Diagnostic Metrics (PDM) destination 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].

2 Performance and Diagnostic Metrics Destination Option

2.1 Destination Options Header

The IPv6 Destination Options Header is used to carry optional information that need 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 RFC2460 [RFC2460].

The IPv6 Performance and Diagnostic Metrics Destination Option (PDM) is an implementation of the IPv6 Destination Options extension header (Next Header value = 60). The PDM does not require time synchronization.

2.2 Performance and Diagnostic Metrics Destination Option

The Performance and Diagnostic Metrics Destination Option (PDM) contains the following fields:

```
TIMEBASE : Base timer unit
SCALEDL  : Scale for Delta Last Received
SCALEDS  : Scale for Delta Last Sent
PSNTP    : Packet Sequence Number This Packet
PSNLR    : Packet Sequence Number Last Received
DELTALR  : Delta Last Received
DELTALS  : Delta Last Sent
```

For a full explanation of the SCALEDL, SCALEDS, DELTALR, DELTALS fields, see [ELK-IPPM].

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

```
0                   1                   2                   3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Option Type  | Option Length |TB |ScaleDL      |   ScaleDS   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   PSN This Packet             |  PSN Last Received            |
|-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Delta Last Received         |  Delta Last Sent              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Option Type

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

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 16.

Time Base

2-bit unsigned integer. It will indicate the unit measurement for this device. That is, for a value of 11 in the Time Base field, a value of 1 in the DELTA fields indicates this device has incremented the time by 1 picosecond. Similarly, for a value of 01 in the Time

Base field, a DELTA value of 1 indicates an increment of 1 microsecond.

The possible values of Time Base are as follows:

00 - milliseconds
01 - microseconds
10 - nanoseconds
11 - picoseconds

Packet Sequence Number This Packet (PSNTP)

16-bit unsigned integer. This field will wrap. It is intended for human use.

Initialized at a random number and monotonically incremented for packet on the 5-tuple. The 5-tuple consists of the source and destination IP addresses, the source and destination ports, and the upper layer protocol (ex. TCP, ICMP, etc).

Operating systems MUST implement a separate packet sequence number counter per 5-tuple. Operating systems MUST NOT implement a single counter for all connections.

Note: This is consistent with the current implementation of the IPID field in IPv4 for many, but not all, stacks.

Packet Sequence Number Last Received (PSNLR)

16-bit unsigned integer. This is the PSN of the packet last received on the 5-tuple.

Scale Delta Last Received (SCALEDLR)

7-bit signed integer. This is the scaling value for the Delta Last Received (DELTALR) field. The possible values are from -64 to +63.

Scale Delta Last Sent (SCALEDLS)

7-bit signed integer. This is the scaling value for the Delta Last Sent (DELTALS) field. The possible values are from -64 to +63.

Delta Last Received (DELTALR)

A 16-bit unsigned integer field.

DELTALR = Send time packet 2 - Receive time packet 1

The value is according to the scale in SCALEDLR.

Delta Last Sent (DELTALS)

A 16-bit unsigned integer field.

Delta Last Sent = Receive time packet 2 - Send time packet 1

The value is in according to the scale in SCALEDLS.

Option Type

The two highest-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. This indicates the following processing requirements:

00 - skip over this option and continue processing the header.

RFC2460 [RFC2460] defines other values for the Option Type field. These MUST NOT be used in the PDM. The other values are as follows:

01 - discard the packet.

10 - discard the packet and, regardless of whether or not the packet's Destination Address was a multicast address, send an ICMP Parameter Problem, Code 2, message to the packet's Source Address, pointing to the unrecognized Option Type.

11 - discard the packet and, only if the packet's Destination Address was not a multicast address, send an ICMP Parameter Problem, Code 2, message to the packet's Source Address, pointing to the unrecognized Option Type.

In keeping with RFC2460 [RFC2460], the third-highest-order bit of the Option Type specifies whether or not the Option Data of that option can change en-route to the packet's final destination.

In the PDM, the value of the third-highest-order bit MUST be 0. The possible values are as follows:

0 - Option Data does not change en-route

1 - Option Data may change en-route

The three high-order bits described above are to be treated as part of the Option Type, not independent of the Option Type. That is, a particular option is identified by a full 8-bit Option Type, not just the low-order 5 bits of an Option Type.

2.3 Time Base

This specification allows for the fact that different CPU TOD clocks use different binary points. For some clocks, a value of 1 could indicate 1 microsecond, whereas other clocks could use the value 1 to indicate 1 millisecond. In the former case, the binary digits to the right of that binary point measure 2**(-n) microseconds, and in the latter case, 2**(-n) milliseconds.

The Time Base allows us to ensure we have a common reference, at the very least, common knowledge of what the binary point is for the transmitted values.

2.4 Timer-value scaling

As discussed in [TRAM-TCPM] we propose storing not an entire time-interval value, but just the most significant bits of that value, along with a scaling factor to indicate the magnitude of the time-interval value. In our case, we will use the high-order 16 bits. The scaling value will be the number of bits in the timer register to the right of the 16th significant bit. That is, if the timer register contains this binary value:

```
1110100011010100101001010001000000000000
<-16 bits     -><-24 bits             ->
```

then, the values stored would be 1110 1000 1101 0100 in binary (E8D4 hexadecimal) for the time value and 24 for the scaling value. Note that the displayed value is the binary equivalent of 1 second expressed in picoseconds.

The following table represents a device which has a TimeBase of picosecond (or 11). For this Time Base value the smallest and simplest time value to represent is 1 picosecond; the encoded value is 1, and the scaling value is 0. Using time values in the first column in the table below, we create the following encoded values and scaling values:

Delta time	Time value in picoseconds	Encoded value	Scaling decimal
1 picosecond	1	0001	0
1 nanosecond	3e8	03e8	0
1 microsecond	f4240	f424	4
1 millisecond	3b9aca00	3b9a	16
1 second	e8d4a51000	e8d4	24
1 minute	3691d6afc000	3691	32
1 hour	cca2e51310000	cca2	36
1 day	132f4579c980000	132f	44
365 days	1b5a660ea44b80000	1b5a	52

Sample binary values (high order 16 bits taken)

```
1 psec         1                                              0001
1 nsec       3e8                                    0011 1110 1000
1 usec     f4240                          1111 0100 0010 0100 0000
1 msec  3b9aca00           0011 1011 1001 1010 1100 1010 0000 0000
1 sec e8d4a51000 1110 1000 1101 0100 1010 0101 0001 0000 0000 0000
```

The need for a signed scaling value is more apparent when the Time Base is lower. If you specify your Time Base is microseconds (or 01) then these tables looks very similar:

Delta time	Time value in microseconds	Encoded value	Scaling decimal
1 microsecond	1	0001	0
1 millisecond	3e8	03e8	0
1 second	f4240	f424	4
1 minute	3938700	e4e1	10
1 hour	d693a400	d693	16
1 day	141dd76000	a0ee	21

An issue arises when the binary point is at the microsecond level, as it is here, but the time differential to be expressed is some number of nanoseconds or picoseconds. For example 1 nanosecond is .001 microseconds, or about .0000000001000001100010010001 in binary. To encode this value we would follow the same procedure:

```
<-       25 bits       ->    Scaling value: -25
0000000001000001100010010001
         <-16 bits     ->    Encoded value: 8312
```

So, the encoded value would be 8312, with a scaling value of -25 which, in the signed 7-bit SCALEDS or SCALEDR field would be represented as 1100111 in binary.

Similarly, if the Time Base is 10, indicating the clock is counting nanoseconds, the encoded value and scaling values for 1 picosecond, or .001 nanoseconds are the same, 8312 and -25, because we've changed the Time Base.

For further discussion on timing and scaling, please see "IPPM Considerations for the IPv6 PDM Destination Option" [ELK-IPPM].

2.5 Header Placement

The PDM destination option MUST be placed as follows:

- Before the upper-layer header. That is, this is the last extension header.

This follows the order defined in RFC2460 [RFC2460]

IPv6 header

Hop-by-Hop Options header

Destination Options header

Routing header

Fragment header

Authentication header

Encapsulating Security Payload header

Destination Options header

upper-layer header

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

The inclusion of a PDM in a packet affects the receiving node's processing of only this single packet. No state is created or modified in the receiving node as a result of receiving a PDM in a packet.

2.6 Implementation Considerations

The PDM destination options extension header SHOULD be turned on by each stack on a host node.

2.6.1 Dynamic Configuration Options

If implemented, each operating system MUST have a default configuration parameter, e.g. diag_header_sys_default_value=yes/no. The operating system MAY also have a dynamic configuration option to change the configuration setting as needed.

If the 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.

The PDM MUST NOT be changed dynamically via packet flow as this may create potential security violation or DoS attack by numerous packets turning the header on and off.

As with all other destination options extension headers, the PDM is for destination nodes only. As specified above, intermediate devices MUST neither set nor modify this field.

2.6.2 Data Length Filtering

Different results for derived metrics, such as, server delay, will be obtained if calculations are done including or excluding packets which have a data length of 0 or 1. Some protocols, for example, TCP, provide acknowledgements which have a length of 0. Keep-alive packets have a data length of 0 or 1.

Operating systems may provide the user a choice of whether to include or exclude packets with a 0 or 1 byte data length.

2.6.3 5-tuple Aging

Within the operating system, metrics must be kept on a 5-tuple basis. The 5-tuple is:

```
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)
```

The question comes of when to stop keeping data or restarting the numbering for a 5-tuple. For example, in the case of TCP, at some point, the connection will terminate. Keeping data in control blocks forever, will have unfortunate consequences for the operating system.

So, the recommendation is to use a known aging parameter such as Max Segment Lifetime (MSL) as defined in Transmission Control Protocol [RFC0793]. The choice of aging parameter is left up to the implementation.

3 Backward Compatibility

The scheme proposed in this document is backward compatible with all the currently defined IPv6 extension headers. According to RFC2460 [RFC2460], if the destination node does not recognize this option, it should skip over this option and continue processing the header.

4 Security Considerations

The PDM MUST NOT be changed dynamically via packet flow as this creates a possibility for potential security violations or DoS attacks by numerous packets turning the header on and off.

5 IANA Considerations

An option type must be assigned by IANA for the Performance and Diagnostic Metrics (PDM) destination option.

6 References

6.1 Normative References

[RFC0793] Postel, J., "Transmission Control Protocol", STD 7, RFC 793, September 1981.

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

[RFC2460] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", RFC 2460, December 1998.

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

6.2 Informative References

[TRAM-TCPM] Trammel, B., "Encoding of Time Intervals for the TCP Timestamp Option-01", Internet Draft, July 2013. [Work in Progress]

[ELK-IPPM] Elkins, N., "IPPM Considerations for the IPv6 PDM Destination Option-01", Internet Draft, October 2014.

7 Acknowledgments

The authors would like to thank Keven Haining, Bill Jouris, Sigfrido Perdomo and Fred Baker for their comments.

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

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

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@bcbsmi.com
http://www.bcbsmi.com