IPv6 Minimum Path MTU Hop-by-Hop Option

<draft-hinden-6man-mtu-option-02>

Bob Hinden
Gorry Fairhurst

IETF105 Montreal
Background

- Current RFC8201 PMTUD isn’t working well.

- This hop-by-hop option came from the idea that it will be more reliable for the Destination to send Path MTU feedback to the Source.
  - Better trust relationship than RFC8201 PMTUD.

- It may not work in all places [RF7872] etc., but we suggest it can help some places.
Changes Since IETF104

- draft-hinden-6man-mtu-option-02 (2019-July-5)
  - Changed option format to also include the Returned MTU value and Return flag and made related text changes in Section 6.2 to describe this behavior.
  - ICMP Packet Too Big messages are no longer used for feedback to the Source host.
  - Added to Acknowledgements Section that a similar mechanism was proposed for IPv4 in 1988 in [RFC1063].
  - Editorial changes.
## New Version of HBH Option

<table>
<thead>
<tr>
<th>Option Type</th>
<th>Option Data Len</th>
<th>Option Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>BB</td>
<td>00</td>
<td>Skip over this option and continue processing.</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>Option data can change en-route to the packet's final destination.</td>
</tr>
<tr>
<td>TTTT</td>
<td>1110</td>
<td>Experimental Option Type from [IANA-HBH].</td>
</tr>
</tbody>
</table>

**Length:** 4 Note the size of the each value field in Option Data field supports Path MTU values from 0 to 65,535 octets.

**Min-PMTU:** n 16-bits. The minimum PMTU in octets, reflecting the smallest link MTU that the packet experienced across the path. This is called the Reported PMTU. A value less than the IPv6 minimum link MTU [RFC8200] should be ignored.

**Rtn-PMTU:** n 15-bits. The returned minimum PMTU, carrying the 15 most significant bits of the latest received Min-PMTU field. The value zero means that no Reported MTU is being returned.

**R** n 1-bit. **R-Flag.** Set by the source to signal that the destination should include the received Reported PMTU in Rtn-PMTU field.
Planned Experiments

- Experiments needed:
  - IPv6 packets with HBH Options (do they make it through the path?)
  - Size of supported PMTU (where is the MTU bottleneck?)
  - What happens in practice (ECMP, etc)

  ... Plan to do look from home, operator and DCs.
  ... More questions will emerge as we do this work!
IETF 105 Hackathon

- Updated VPP Router implementation (Ole Troan)
- Updated Wireshark dissector (Bob Hinden)
  - Modified Wireshark instead of using lua plugin

<table>
<thead>
<tr>
<th>No.</th>
<th>Time</th>
<th>Source</th>
<th>Destination</th>
<th>Protocol</th>
<th>Length</th>
<th>Info</th>
</tr>
</thead>
</table>

- Frame 823: 78 bytes on wire (624 bits), 78 bytes captured (624 bits) on interface 0
  0110 .... = Version: 6
  .... 0000 0000 .... .... .... .... = Traffic Class: 0x00 (DSCP: CS0, ECN: Not-ECT)
  .... .... 0000 0000 0000 0000 0000 = Flow Label: 0x000000
  Payload Length: 16
  Next Header: IPv6 Hop-by-Hop Option (0)
  Hop Limit: 62

- IPv6 Hop-by-Hop Option
  - Next Header: ICMPv6 (58)
  - Length: 0
  - [Length: 8 bytes]
- Path MTU Option (0x3E)
  - Type: Path MTU Option (0x3E) (0x3E)
  - Length: 4
  - Minimum PMTU: 9000

Return PMTU: 18002

Return Flag: True

Internet Control Message Protocol v6

0000 38 f9 d3 40 3c a0 fc 33 42 21 38 08 86 dd 60 00 8- @<= 3 B!8 8...}
0010 00 00 00 10 00 3e 20 01 86 7c 12 38 06 01 14 5a 4- > 0- | 0- -Z
0020 0f fe 00 bd 9d 70 28 01 86 7c 03 70 01 28 00 00 8... | p > | p ({
0030 db 7f 10 db aa 08 3a 00 3e 04 23 28 46 53 00 00 00 00 00 > #FS
0040 2f 88 00 00 00 00 75 ff 97 00 64 c4 f9 a5 /... u #d... |
Next Steps

- Continue experiments (please talk to us).
- W.G. Adopt as working group document
- Ask AD and IANA for early allocation
  - Needed for Internet wide experimentation
QUESTIONS / COMMENTS?
BACKGROUND
Motivation

- PMTUD [RFC8201] doesn’t work well in the Internet
  - Nodes in the middle of the network may not send a ICMP Packet Too Big message.
  - Path often doesn’t/can’t return the PTB message to a sender.
  - Nodes mostly rely on MSS for TCP and default to 1280 for UDP.

- Problematic for transport encapsulations and tunneling that reduce available MTU.

- Limits usefulness of 10G and 100G Ethernet.
  - 1280 octet packets need 977K pps at 10G.
  - 9000 octet packets need 129K pps at 10G.
Investigating Approaches to Provide MTU Feedback

- Endpoint PTB message to sent to source
- Reflection of value in a HBH option on the same flow
- Reflection of value within a transport parameter for the flow
Goals

- Learn by testing / experimentation if this provides enough value to justify deployment.

- Understand how to integrate this as a part of a framework that is robust to loss or probes e.g. (D)PLPMTUD.