IPv6 Minimum Path MTU Hop-by-Hop Option

<draft-hinden-6man-mtu-option-01>

Bob Hinden

Gorry Fairhurst

IETF104 Prague

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 IETF103

- draft-hinden-6man-mtu-option-01 (2019-March-11)
  - Changed requested status from Standards Track to Experimental to use an experimental option type (11110). Removed request for IANA Option assignment.
  - Added Section 2 "Motivation and Problem Solved" section to better describe the purpose of this document.
  - Added Appendix A describing planned experiments and how the results will be measured (more next revision).
  - Editorial changes.

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

Planned Experiments

- Experiments needed:
  - How likely is it that an IPv6 H-B-H option will be forwarded to the remote node?
  - How likely is it that a PTB message from a remote node is returned to the source?
  - How easy is this to implement?
  - How much support is there for jumbo frames?

  - … More questions will emerge as we do this work!

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.

IETF 104 Hackathon

- Initial Linux host implementation (Tom Herbert, Vladimir)
- Initial BSD host implementation (Tom Jones)
- VPP Router implementation (Ole Troan)
- P4 router implementation (Luuk Hendriks, Ronald Vanderpol)
- Wireshark dissector (Bob Hinden)

New Version of HBH Option


```
 Option     Option     Option
  Type     Data Len    Data
+--------+--------+--------+--------+-+-------+--------+
|BBCTTTTT|00000100|     value 1      |R|    value 2     |
+--------+--------+--------+--------+-+-------+--------+
```

Option Type:

BB 00 Skip over this option and continue processing.

C 1 Option data can change en route to the packet's final destination.

TTTT 11110 Experimental Option Type from [IANA-HBH].

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

Value 1: n The Reported PMTU in octets, reflecting the smallest link MTU that the packet experienced across the path.

R n Return bit. Set by the source to signal the destination that it should include the received minimum MTU in Value 2.

Value 2: n The retuned minimum MTU from from Value 1 divided by two (shift right one bit).

Wireshark

No.	Time	Source	Destination	Protocol	Length	Info
10	2.003583	3::1111	3::1111	MTUOP…	94	
11	2.003902	3::1111	3::1111	MTUOP…	94	

```
▶ Frame 10: 94 bytes on wire (752 bits), 94 bytes captured (752 bits)
▶ Ethernet II, Src: SuperMic_5f:3c:c5 (00:25:90:5f:3c:c5), Dst: AsustekC_92:bc:ee (70:8b:cd:92:bc:ee)
▶ Internet Protocol Version 6, Src: 3::1111, Dst: 3::1111
▼ MTU Hop by Hop Option Data
    Option Next Header: 17
    Option Extension Length: 0
    Option Type: 0x3e
    Option Payload Length: 4
    Minimum PMTU: 9000
    1... .... .... .... = Respond Flag: True
    Returned Minimum PMTU: 0
```

Next Steps

- Continue experiments (please talk to us).

QUESTIONS / COMMENTS?