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


```
 Option    Option    Option
  Type    Data Len   Data
+--------+--------+--------+--------+-------+--------+-+
|BBCTTTTT|00000100|     Min-PMTU     |   Rtn-PMTU     |R|
+--------+--------+--------+--------+-------+--------+-+
```

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.

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

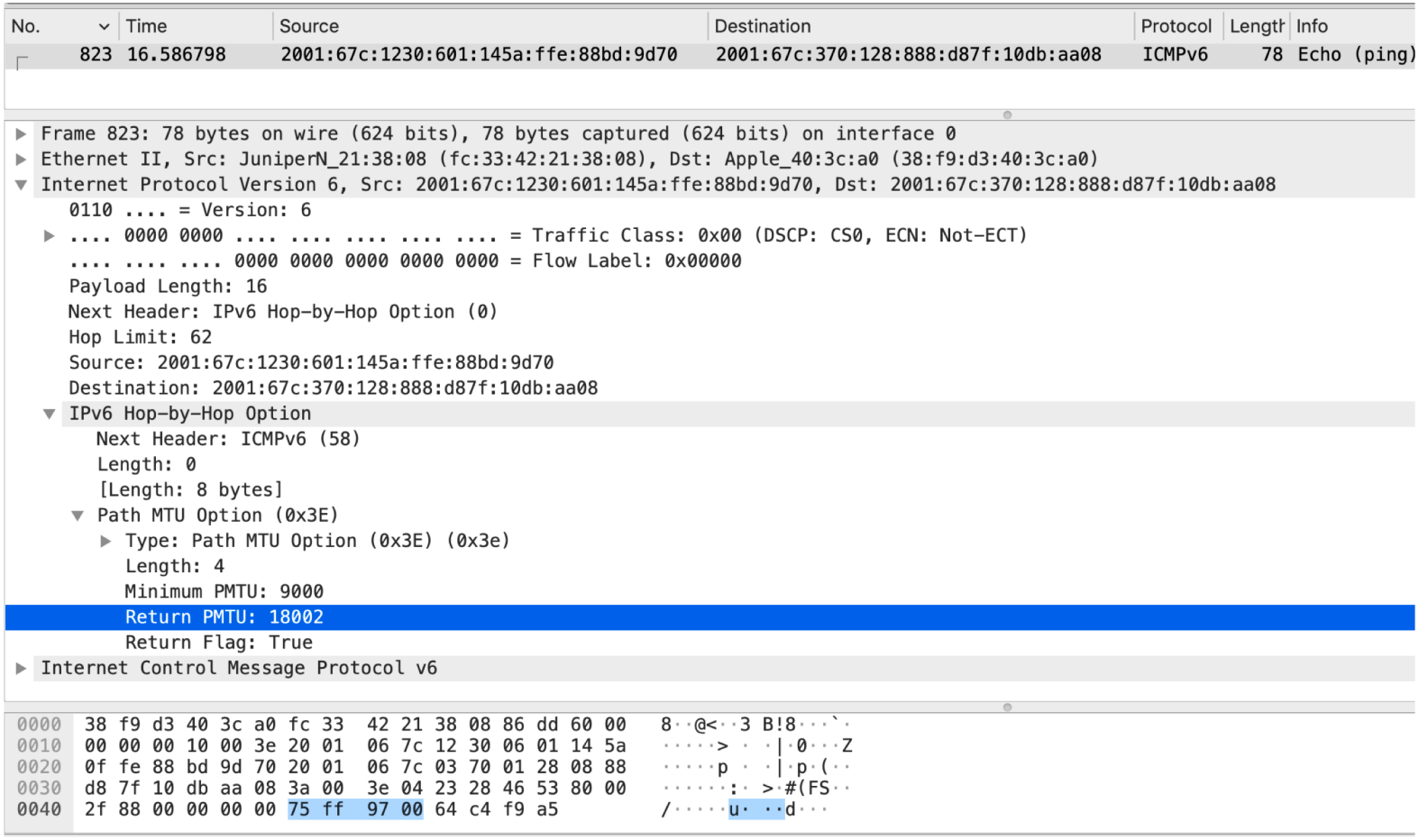
- Experiments needed:

- 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



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.