What should networks do with IPv6 Extension Headers?

Ana Custura
Gorry Fairhurst

* University of Aberdeen with funding from RIPE NCC
IPV6

IPv6 was standardised in the 1990's [RFC2460]

- Became Full Standard in 2017 [RFC 8200]

<table>
<thead>
<tr>
<th>Version</th>
<th>DSCP/ToS</th>
<th>ECN</th>
<th>Flow Label</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Payload Length</th>
<th>Next Header</th>
<th>Hop Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

128 bit Source Address

128 bit Destination Address

<table>
<thead>
<tr>
<th>Next Header</th>
<th>Header Length</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Header Extensions (if any)
IPv6 promises

- Larger Address Space
- More Efficient Forwarding/Routing
- Improved IP Packet Fragmentation*
- Multicast
- End-to-end Security (aka IPSEC)
- Extensibility

Fix to lack of IPv4 address space

RIPE ‘86
IPv6 promises

- Larger Address Space
- More Efficient Forwarding/Routing
- Improved IP Packet Fragmentation*
- Multicast
- End-to-end Security (aka IPSEC)
- Extensibility

Fix to lack of IPv4 address space

Fix to lack of extension in IPv4

*After some refinements

RIPE '86
IPv6 promises

- Larger Address Space
  - Fix to lack of IPv4 address space
- More Efficient Forwarding/Routing
- Improved IP Packet Fragmentation*
  - *After some refinements
- Multicast
- End-to-end Security (aka IPSEC)
- Extensibility
  - Fix to lack of extension in IPv4

Other ways have emerged, such as QUIC

RIPE ‘86
IPv6 promises

- Larger Address Space
- More Efficient Forwarding/Routing
- Improved IP Packet Fragmentation*
- Multicast
- End-to-end Security (aka IPSEC)
- Extensibility

Fix to lack of IPv4 address space

*After some refinements

Other ways have emerged, such as QUIC

This talk!
### Extensibility - EH

**IPv6 Base Header**

<table>
<thead>
<tr>
<th>Protocol Number</th>
<th>Description</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>IPv6 Hop-by-Hop Option</td>
<td>[RFC8200]</td>
</tr>
<tr>
<td>43</td>
<td>Routing Header for IPv6</td>
<td>[RFC8200] [RFC5095]</td>
</tr>
<tr>
<td>44</td>
<td>Fragment Header for IPv6</td>
<td>[RFC8200]</td>
</tr>
<tr>
<td>50</td>
<td>Encapsulating Security Payload</td>
<td>[RFC4303]</td>
</tr>
<tr>
<td>51</td>
<td>Authentication Header</td>
<td>[RFC4302]</td>
</tr>
<tr>
<td>60</td>
<td>Destination Options for IPv6</td>
<td>[RFC8200]</td>
</tr>
<tr>
<td>135</td>
<td>Mobility Header</td>
<td>[RFC6275]</td>
</tr>
<tr>
<td>139</td>
<td>Host Identity Protocol</td>
<td>[RFC7401]</td>
</tr>
<tr>
<td>140</td>
<td>Shim6 Protocol</td>
<td>[RFC5533]</td>
</tr>
<tr>
<td>253,254</td>
<td>Use for experimentation and testing</td>
<td>[RFC3692] [RFC4727]</td>
</tr>
</tbody>
</table>

**Extension Headers**

- [IPv6](#)
- [HBH OPT](#)
- [DOPT](#)
- [DOPT](#)
- [TCP](#)

**Upper layer protocol**

RIPE ‘86

[https://www.iana.org/assignments/ipv6-parameters/ipv6-parameters.xhtml](https://www.iana.org/assignments/ipv6-parameters/ipv6-parameters.xhtml)
# Extensibility - EH

## Extension Headers

<table>
<thead>
<tr>
<th>Protocol Number</th>
<th>Description</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>IPv6 Hop-by-Hop Option</td>
<td>[RFC8200]</td>
</tr>
<tr>
<td>43</td>
<td>Routing Header for IPv6</td>
<td>[RFC8200] [RFC5095]</td>
</tr>
<tr>
<td>44</td>
<td>Fragment Header for IPv6</td>
<td>[RFC8200]</td>
</tr>
<tr>
<td>50</td>
<td>Encapsulating Security Payload</td>
<td>[RFC4303]</td>
</tr>
<tr>
<td>51</td>
<td>Authentication Header</td>
<td>[RFC4302]</td>
</tr>
<tr>
<td>60</td>
<td>Destination Options for IPv6</td>
<td>[RFC8200]</td>
</tr>
<tr>
<td>135</td>
<td>Mobility Header</td>
<td>[RFC6275]</td>
</tr>
<tr>
<td>139</td>
<td>Host Identity Protocol</td>
<td>[RFC7401]</td>
</tr>
<tr>
<td>140</td>
<td>Shim6 Protocol</td>
<td>[RFC5533]</td>
</tr>
<tr>
<td>253,254</td>
<td>Use for experimentation and testing</td>
<td>[RFC3692] [RFC4727]</td>
</tr>
</tbody>
</table>

IPv6 Base Header

Some EHs carry ‘Options’

[https://www.iana.org/assignments/ipv6-parameters/ipv6-parameters.xhtml](https://www.iana.org/assignments/ipv6-parameters/ipv6-parameters.xhtml)

RIPE ‘86
EH concerns in RFC 9098 (2021)

- Slow-path processing of EHs
- Buggy implementations* -> DoS
- Complexity not bounded: can reduce router forwarding rate
- Large EH can exceed router parsing buffer

* To this date, vulnerabilities still found: https://www.interruptlabs.co.uk/articles/linux-ipv6-route-of-death
EH concerns in RFC 9098 (2021)

- Slow-path processing of EHs
- Buggy implementations* -> DoS
- Complexity not bounded: can reduce router forwarding rate
- Large EH can exceed router parsing buffer

Some EHs had a rocky start

Measurements in RFC 7872 show many networks drop packets with EH

* To this date, vulnerabilities still found: https://www.interruptlabs.co.uk/articles/linux-ipv6-route-of-death
Renewed Interest in EHs

- IPv6 Segment Routing type (SRv6) [RFC8986]
- Service Management and Performance Measurement using PDM [RFC8250]
- In-situ Operations, Administration, and Maintenance [RFC9378]
- AltMark Measurement DO and HbH Options [RFC9343]
- minPMTU HBH Option [RFC9268]

**ASICs are emerging that can process EHs at line speed!**
Renewed Interest in EHs

- IPv6 Segment Routing type (SRv6) [RFC8986]
- Service Management and Performance Measurement using PDM [RFC8250]
- In-situ Operations, Administration, and Maintenance [RFC9378]
- AltMark Measurement DO and HbH Options [RFC9343]
- minPMTU HBH Option [RFC9268]

Can Options be used more widely in the Internet?

RIPE ‘86
Existing Measurements

- Focus on **Destination Options (DOPT)** and **Hop-by-Hop Options (HBHOPT) EHs**

- Let's measure survival of packets with EH

<table>
<thead>
<tr>
<th></th>
<th>Destination Option EH</th>
<th>Hop-by-Hop Option EH</th>
</tr>
</thead>
<tbody>
<tr>
<td>RFC 7872 (2016) [1] - server edge</td>
<td>80-90%</td>
<td>45-60%</td>
</tr>
<tr>
<td>My own (2018) data [2] - server edge</td>
<td>70-75%</td>
<td>15-20%</td>
</tr>
<tr>
<td>APNIC (2022) [3] - client edge</td>
<td>30-80%</td>
<td>0%</td>
</tr>
<tr>
<td>JAMES (2022) [4] - core</td>
<td>94-97%</td>
<td>8-9%</td>
</tr>
</tbody>
</table>
Experiment 1: Survival

- ~5500 IPv6-enabled probes in RIPE, globally distributed
- Testing survival by sending packets to 7 targets (UK, US, Canada, Australia, Zambia, Kazakhstan, France)
  - \{TCP, UDP\} to port 443
    - \{DOPT, HBHOPT\} + control IPv6 packets
  - Survives if packet reaches destination AS
Survival at a Glance

DOPTs

- 8B PadN option
- High survival for **DOPTs**
- Difference between TCP and UDP

![Diagram showing network traffic and survival rates between DOPT and TCP/UDP]](image-url)

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Survival Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCP</td>
<td>~68%</td>
</tr>
<tr>
<td>UDP</td>
<td>~92%</td>
</tr>
</tbody>
</table>

RIPE ‘86
Survival at a Glance

HBHOPTs

- 8B PadN option
- HBHOPTs survive some paths
- Difference between TCP and UDP

\[ \text{5000 probes} \quad \text{Source} \quad \text{Hop 1} \quad \text{Hop 2} \quad \text{Hop 3} \quad \ldots \quad \text{Hop N} \quad \text{Destinations} \]

- HBHOPT
  - \( \sim 11\% \) UDP
  - \( \sim 9\% \) TCP

RIPE '86
### Per-AS Survival (UK path)

#### DOPT

The **local AS** is responsible for most of the drops:
- 5% for UDP
- 25% for TCP

<table>
<thead>
<tr>
<th></th>
<th>1st AS</th>
<th>AS1&gt;AS2</th>
<th>∞</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DOPT UDP 8B</strong></td>
<td>95.3%</td>
<td>93%</td>
<td>91.5%</td>
</tr>
<tr>
<td><strong>DOPT TCP 8B</strong></td>
<td>74.7%</td>
<td>70%</td>
<td>68.5%</td>
</tr>
</tbody>
</table>

#### HBHOPT

The **local AS** is responsible for most of the drops:
- 68% for UDP
- 74% for TCP

<table>
<thead>
<tr>
<th></th>
<th>1st AS</th>
<th>AS1&gt;AS2</th>
<th>2nd AS</th>
<th>AS2&gt;AS3</th>
<th>∞</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HBHOPT UDP 8B</strong></td>
<td>31.4%</td>
<td>20.1%</td>
<td>15%</td>
<td>12.2%</td>
<td>11.4%</td>
</tr>
<tr>
<td><strong>HBHOPT TCP 8B</strong></td>
<td>26.9%</td>
<td>16.3%</td>
<td>13.9%</td>
<td>9.7%</td>
<td>8.6%</td>
</tr>
</tbody>
</table>

Drops are considered to be within the AS if the next hop on a control measurement is also in that AS. If the next hop would otherwise be in a different AS, then the drop is attributed to the AS boundary.

RIPE '86
What if packets would traverse the first AS?

- Most probes have public IPv6 addresses
- Reverse traceroute on paths where drops happen in first AS
- Same protocol/port
- Does the packet reach original AS?

RIPE ‘86
What if packets would traverse the first AS?

**DOPTs**
Reverse traceroute on paths with drops in first AS (n=271 paths for UDP):
95 - 97% make it back to the original AS.

<table>
<thead>
<tr>
<th>%predicted traversal</th>
<th>DOPT UDP (UK)</th>
<th>DOPT UDP (Canada)</th>
</tr>
</thead>
<tbody>
<tr>
<td>~96%</td>
<td>~96%</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>%predicted traversal</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>~17%</td>
<td>60% packets get dropped at LINX peering</td>
</tr>
<tr>
<td>~25%</td>
<td></td>
</tr>
</tbody>
</table>

**HBHOPTs**
Reverse traceroute on paths with drops in first AS (n=3150 paths for UDP):
10 - 17% make it back to the original AS.

Transit networks drop more packets with HBHOPTs
Why?

- Network/Firewall policy (e.g. Fastly)
- Different router designs
- Different devices (CPE, load balancers, firewalls, IDS) wanting access to upper layer protocols
- End-systems (NICs that do processing in hosts)
  - Is EH size a factor? Is full chain size a factor?
Experiment 2: Size

- \{TCP, UDP\} to port 443
- \{DOPT, HBHOPT\} + control measurement
- \{8,16,32,40,48,56,64\} B in size to one target
- Survival is successful if packet reaches destination AS
Traversal vs Size

- TCP sees the biggest drop in traversal at 48B: $48 + 20 = 68B$ (108B total)
- UDP sees the biggest drop at 56B: $56 + 8 = 64B$ (104B total)
- Is this due to EH size or IPv6 total chain size?
- 40B is the max for IPv4 options
TCP sees the biggest drop in traversal at 48B: $48 + 20 = 68B$ (108B total)

UDP sees the biggest drop at 56B: $56 + 8 = 64B$ (104B total)

Is this due to EH size or IPv6 total chain size?

40B is the max for IPv4 options

Where EHs can be used, 40B often works
Experiment 3: ECMP

- ECMP uses header information for load-balancing
- UDP to port 443 from ~850 probes
  - \{DOPT, HBHOPT\} + control measurement
  - We measure 16 Paris ID variations to the same target (Flow Label + source port combinations)
• Not all devices are equipped to handle flows that mix packets with and without EHs

• Motivates the use of Flow Label for ECMP
No EH measurement

Same source-destination pair measured with Destination Options EH Packets

RIPE '86
Same source-destination pair measured with **Hop-by-Hop Options** EH packets
What should networks do with them?

- Firewall, firewall … but only if you need to!
- IPv6 is being extended within domains
- Unnecessary barriers bad for innovation
- More capable ASICs - > Forwarding + processing without impacting performance

3 new IETF drafts might help: draft-ietf-6man-eh-limits, draft-ietf-6man-hbh-processing, draft-ietf-v6ops-hbh
What next?

- Fragmentation got ‘fixed’ after trials and tribulations

- What about Options:

  ...within a domain? It is low-risk, can be and IS done now

  ...opportunistically in the Internet? DOPTs almost there

What about in 5 years’ time?
References

- [3] https://blog.apnic.net/2022/10/13/ipv6-extension-headers-revisited/