Workgroup: IPv6 Operations
Internet-Draft: draft-vyncke-v6ops-james-01
Published: 20 March 2022
Intended Status: Informational
Expires: 21 September 2022
Authors: É. Vyncke R. Léas J. Iurman
Cisco Université de Liège Université de Liège

Just Another Measurement of Extension header Survivability (JAMES)

Abstract

In 2016, RFC7872 has measured the drop of packets with IPv6 extension headers. This document presents a slightly different methodology with more recent results. It is still work in progress.

About This Document

This note is to be removed before publishing as an RFC.

The latest revision of this draft can be found at https://evyncke.github.io/v6ops-james/draft-vyncke-v6ops-james.html. Status information for this document may be found at https://datatracker.ietf.org/doc/draft-vyncke-v6ops-james/.

Discussion of this document takes place on the IPv6 Operations Working Group mailing list (mailto:v6ops@ietf.org), which is archived at https://mailarchive.ietf.org/arch/browse/v6ops/.

Source for this draft and an issue tracker can be found at https://github.com/evyncke/v6ops-james.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at https://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on 21 September 2022.

Table of Contents

1. Introduction

In 2016, [RFC7872] has measured the drop of packets with IPv6 extension headers on their transit over the global Internet. This document presents a slightly different methodology with more recent results. Since then, [I-D.draft-ietf-opsec-ipv6-eh-filtering] has provided some recommendations for filtering transit traffic, so there may be some changes in providers policies.

It is still work in progress, but the authors wanted to present some results at IETF-113 (March 2022). The code is open source and is available at [GITHUB].

2. Methodology

In a first phase, the measurement is done between collaborating IPv6 nodes, a.k.a. vantage points, spread over the Internet and multiple Autonomous Systems (ASs). As seen in Section 3.2, the source/ destination/transit ASs include some "tier-1" providers per [TIER1], so, they are probably representative of the global Internet core.

Relying on collaborating nodes has some benefits:

* propagation can be measured even in the absence of any ICMP message or reply generated by the destination;

* traffic timing can be measured accurately to answer whether extension headers are slower than plain IP6 packets;

* traffic can be captured into .pcap [I-D.draft-ietf-opsawg-pcap] file at the source and at the destination for later analysis.

Future phases will send probes to non-collaborating nodes with a much reduced probing speed. The destination will include [ALEXA] top-n websites, popular CDN, as well as random prefix from the IPv6 global routing table. A revision of this IETF draft will describe those experiments.

3. Measurements

3.1. Vantage Points

Several servers were used worldwide (albeit missing Africa and China as the authors were unable to find IPv6 servers in these regions). Table 1 lists all the vantage points together with their AS number and country.

ASN	AS Name	Country code	Location
7195	Edge Uno	AG	Buenos Aires
12414	NL-SOLCON SOLCON	NL	Amsterdam
14061	Digital Ocean	CA	Toronto, ON
14061	Digital Ocean	USA	New York City, NY
14601	Digital Ocean	DE	Francfort
14601	Digital Ocean	IN	Bangalore
14601	Digital Ocean	SG	Singapore
16276	OVH	AU	Sydney
16276	OVH	PL	Warsaw

ASN	AS Name	Country code	Location
44684	Mythic Beasts	UK	Cambridge
47853	Hostinger	US	Ashville, NC
60011	MYTHIC-BEASTS-USA	US	Fremont, CA
198644	GO6	SI	Ljubljana

Table 1: All vantage AS

3.2. Tested Autonomous Systems

During first phase (traffic among fully-meshed collaborative nodes), Table 2 show the ASs for which our probes have collected data.

AS Number	AS Description	Comment
174	COGENT-174, US	Tier-1
1299	TWELVE99 Twelve99, Telia Carrier, SE	Tier-1
2914	NTT-COMMUNICATIONS-2914, US	Tier-1
3320	DTAG Internet service provider operations, DE	Tier-1
3356	LEVEL3, US	Tier-1
4637	ASN-TELSTRA-GLOBAL Telstra Global, HK	Regional Tier
4755	TATACOMM-AS TATA Communications formerly VSNL is Leading ISP, IN	
5603	SIOL-NET Telekom Slovenije d.d., SI	
6453	Tata Communication	Tier-1
6762	SEABONE-NET TELECOM ITALIA SPARKLE S.p.A., IT	Tier-1
6939	HURRICANE, US	Regional Tier
7195	EDGEUNO SAS, CO	
8447	A1TELEKOM-AT A1 Telekom Austria AG, AT	
9498	BBIL-AP BHARTI Airtel Ltd., IN	
12414	NL-SOLCON SOLCON, NL	
14061	DIGITALOCEAN-ASN, US	
16276	OVH, FR	
21283	A1SI-AS A1 Slovenija, SI	
34779	T-2-AS AS set propagated by T-2 d.o.o., SI	
44684	MYTHIC Mythic Beasts Ltd, GB	
60011	MYTHIC-BEASTS-USA, GB	
198644	GO6, SI	

Table 2: All AS (source/destination/transit)

The table attributes some tier qualification to some ASs based on the Wikipedia page [TIER1], but there is no common way to decide who is a tier-1. Based on some CAIDA research, all the above (except GO6, which is a stub network) are transit providers.

While this document lists some operators, the intent is not to build a wall of fame or a wall of shame but more to get an idea about which kind of providers drop packets with extension headers and how widespread the drop policy is enforced and where, i.e., in the access provider or in the core of the Internet.

3.2.1. Drop attribution to AS

Comparing the traceroutes with and without extension headers allows the attribution of a packet drop to one AS. But, this is not an easy task as inter-AS links often use IPv6 address of only one AS (if not using link-local per [RFC7704]). This document uses the following algorithm to attribute the drop to one AS for packet sourced in one AS and then having a path traversing AS#foo just before AS#bar:

* if the packet drop happens at the first router (i.e., hop limit == 1 does not trigger an ICMP hop-limit exceeded), then the drop is assumed to this AS as it is probably an ingress filter on the first router (i.e., the hosting provider in most of the cases - except if collocated with an IXP).

* if the packet drop happens in AS#foo after one or more hop(s) in AS#bar, then the drop is assumed to be in AS#foo ingress filter on a router with an interface address in AS#foo

* if the packet drop happens in AS#bar after one or more hop(s) in AS#bar before going to AS#foo, then the drop is assumed to be in AS#foo ingress filter on a router with an interface address in AS#bar

In several cases, the above algorithm was not possible (e.g., some intermediate routers do not generate an ICMP unreachable hop limit exceeded even in the absence of any extension headers), then the drop is not attributed. Please also note that the goal of this document is not to 'point fingers to operators' but more to evaluate the potential impact. I.e., a tier-1 provider dropping packets with extension headers has a much bigger impact on the Internet traffic than an access provider.

Future revision of this document will use the work of [MLAT_PEERING].

3.3. Tested Extension Headers

In the first phase among collaborating vantage points, packets always contained either a UDP payload or a TCP payload, the latter is sent with only the SYN flag set and with data as permitted by section 3.4 of [RFC793] (2nd paragraph). A usual traceroute is done with only the UDP/TCP payload without any extension header with varying hop-limit in order to learn the traversed routers and ASs.

Then, several UDP/TCP probes are sent with a set of extension headers:

- hop-by-hop and destination options header containing:
  - one PadN option for an extension header length of 8 octets,
  - one unknown option with the "discard" bits for an extension header length of 8 octets,
  - multiple PadN options for an extension header length of 256 octets,
  - one unknown option (two sets with "discard" and "skip" bits) for the destination options header length of 16, 32, 64, and 128 octets,
  - one unknown option (two sets with "discard" and "skip" bits) for an extension header length of 256 and 512 octets.
- routing header with routing types from 0 to 6 inclusive;
- atomic fragment header (i.e., M-flag = 0 and offset = 0) of varying frame length 512, 1280, and 1500 octets;
- non-atomic first fragment header (i.e., M-flag = 1 and offset = 0) of varying frame length 512, 1280, and 1500 octets;
- authentication header with dummy SPI followed by UDP/TCP header and a 38 octets payload.

In the above, length is the length of the extension header itself except for the fragmentation header where the length is the IP packet length (i.e., including the IPv6, and TCP/UDP headers + payload).

For hop-by-hop and destination options headers, when required multiple PadN options were used in order to bypass some Linux kernels that consider a PadN larger than 8 bytes is an attack, see section 5.3 of [BCP220], even if multiple PadN options violates section 2.1.9.5 of [RFC4942].

In addition to the above extension headers, other probes were sent with next header field of IPv6 header set to:

- 59, which is "no next header", especially whether extra octets after the no next header as section 4.7 [RFC8200] requires that "those octets must be ignored and passed on unchanged if the packet is forwarded";

*143, which is Ethernet payload (see section 10.1 of [RFC8986]).

4. Results

This section presents the current results out of phase 1 (collaborating vantage points) testing. About 4860 experiments were run, one experiment being defined by sending packets between 2 vantage points with hop-limit varying from 1 to the number of hops between the two vantage points and for all the extension headers described in Section 3.3.

4.1. Routing Header

Table 3 lists all routing header types and the percentage of experiments that were successful, i.e., packets with routing header reaching their destination:

Routing Header Type	%-age of packets reaching destination
0	80.9%
1	99.5%
2	99.5%
3	99.5%
4	69.0%
5	99.5%
6	99.3%

Table 3: Per Routing Header Types Transmission

Table 4 lists the few ASs that drop packets with the routing header type 0 (the original source routing header, which is now deprecated).

AS Number	AS description
6939	HURRICANE, US

Table 4: AS Dropping Routing Header Type 0

It is possibly due to a strict implementation of [RFC5095] but it is expected that no packet with routing header type 0 would be transmitted anymore. So, this is not surprising.

Table 5 lists the few ASs that drop packets with the routing header type 4 (Segment Routing Header [RFC8754]).

AS Number	AS description
16276	OVH, FR

Table 5: AS Dropping
Routing Header Type 0

This drop of SRH was to be expected as SRv6 is specified to run only in a limited domain.

Other routing header types (1 == deprecated NIMROD [RFC1753], 2 == mobile IPv6 [RFC6275], 3 == RPL [RFC6554], and even 5 == CRH-16 and 6 == CRH-32[I-D.draft-bonica-6man-comp-rtg-hdr]) can be transmitted over the global Internet without being dropped (assuming that the 0.5% of dropped packets are within the measurement error).

4.2. Hop-by-Hop Options Header

Many ASs drop packets containing either hop-by-hop options headers per Table 6 below:

Option Type	Length	%-age of packets reaching destination
Skip	8	5.8%
Discard	8	0.0%
Skip one large PadN	256	1.9%
Skip multiple PadN	256	0.0%
Discard	256	0.0%
Skip	512	1.9%
Discard	512	0.0%

Table 6: Hop-by-hop Transmission

It appears that hop-by-hop options headers cannot reliably traverse the global Internet; only small headers with 'skipable' options have some chances. If the unknown hop-by-hop option has the 'discard' bits, it is dropped per specification.

4.3. Destination Options Header

Many ASs drop packets containing destination options headers per Table 7:

Length	Multiple PadN	%-age of packets reaching destination
8	No	99.3%
16	No	99.3%
32	No	93.3%
64	No	41.6%
128	No	4.5%
256	No	2.6%
256	Yes	2.6%
512	No	2.6%

Table 7: Hop-by-hop Transmission

The measurement did not find any impact of the discard/skip bits in the destination headers options, probably because the routers do not look inside the extension headers into the options. The use of a single large PadN or multiple 8-octet PadN options does not influence the result.

The size of the destination options header has a major impact on the drop probability. It appears that extension header larger than 16 octets already causes major drops. It may be because the 40 octets of the IPv6 header + the 16 octets of the extension header (total 56 octets) is still below some router hardware lookup mechanisms while the next measured size (extension header size of 64 octets for a total of 104 octets) is beyond the hardware limit and some AS has a policy to drop packets where the TCP/UDP ports are unknown...

4.4. Fragmentation Header

The propagation of two kinds of fragmentation headers was analysed: atomic fragment (offset == 0 and M-flag == 0) and plain first fragment (offset == 0 and M-flag == 1). The Table 8 displays the propagation differences.

M-flag	%-age of packets reaching destination
0 (atomic)	70.2%
1	99.0%

Table 8: IPv6 Fragments Transmission

The size of the overall IP packets (512, 1280, and 1500 octets) does not have any impact on the propagation.

4.5. No extension headers drop at all

Table 9 lists some ASs that do not drop transit traffic (except for routing header type 0) and follow the recommendations of [I-D.draft-ietf-opsec-ipv6-eh-filtering]. This list includes tier-1 transit providers (using the "regional" tag per [TIER1]):

AS Number	AS Description	Comment
4637	ASN-TELSTRA-GLOBAL Telstra Global, HK	Regional Tier
4755	TATACOMM-AS TATA Communications formerly VSNL is Leading ISP, IN	
21283	A1SI-AS A1 Slovenija, SI	
60011	MYTHIC-BEASTS-USA, GB	

Table 9: ASs Not Dropping Packets with Extension Headers

Some ASs also drop only large (more than 8 octets) destination options headers, see Table 10:

AS Number	AS Description	Largest Forwarded Dest.Opt. Size
6453	Tata Communication	8
1299	TWELVE99 Twelve99, Telia Carrier, SE	8
174	COGENT-174, US	8

Table 10: ASs Only Dropping Packets with Large Destination Options Headers

4.6. Special Next Headers

Measurements also include two protocol numbers that are mainly new use of IPv6. Table 11 indicates the percentage of packets reaching the destination.

Next Header	%-age of packets reaching destination
NoNextHeader (59)	99.7%
Ethernet (143)	99.2%

Table 11: Transmission of Special IP Protocols

The above indicates that those IP protocols can be transmitted over the global Internet without being dropped (assuming that the 0.3-0.8% of dropped packets are within the measurement error).

5. Summary of the collaborating parties measurements

While the analysis has areas of improvement (geographical distribution and impact on latency), it appears that:

- authentication and non-atomic fragmentation headers can traverse the Internet;
- only routing headers types 0 and 4 experiment problems over the Internet, other types have no problems;
- hop-by-hop options headers do not traverse the Internet;
- destination options headers are not reliable enough when it exceeds 16 octets.

Of course, the next phase of measurement with non-collaborating parties will probably give another view.

6. Security Considerations

While active probing of the Internet may be considered as an attack, this measurement was done among collaborating parties and using the probe attribution technique described in [I-D.draft-vyncke-opsec-probe-attribution] to allow external parties to identify the source of the probes if required.

7. IANA Considerations

This document has no IANA actions.

8. References

8.1. Normative References

[RFC793] Postel, J., "Transmission Control Protocol", STD 7, RFC 793, DOI 10.17487/RFC0793, September 1981, <https://www.rfc-editor.org/rfc/rfc793>.

[RFC8200] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", STD 86, RFC 8200, DOI 10.17487/RFC8200, July 2017, <https://www.rfc-editor.org/rfc/rfc8200>.

8.2. Informative References

[ALEXA] "The top 500 sites on the web", n.d., <https://www.alexa.com/topsites>.

[BCP220] Chown, T., Loughney, J., and T. Winters, "IPv6 Node Requirements", BCP 220, RFC 8504, January 2019. <https://www.rfc-editor.org/info/bcp220>

[GITHUB] Léas, R., "james", n.d., <https://gitlab.uliege.be/Benoit.Donnet/james>.

[I-D.draft-bonica-6man-comp-rtg-hdr] Bonica, R., Kamite, Y., Alston, A., Henriques, D., and L. Jalil, "The IPv6 Compact Routing Header (CRH)", Work in Progress, Internet-Draft, draft-bonica-6man-comp-rtg-hdr-27, 15 November 2021, <https://datatracker.ietf.org/doc/html/draft-bonica-6man-comp-rtg-hdr-27>.

[I-D.draft-ietf-opsawg-pcap] Harris, G. and M. C. Richardson, "PCAP Capture File Format", Work in Progress, Internet-Draft,

draft-ietf-opsawg-pcap-00, 25 October 2021, <https://datatracker.ietf.org/doc/html/draft-ietf-opsawg-pcap-00>.

[I-D.draft-ietf-opsec-ipv6-eh-filtering] Gont, F. and W. (. Liu, "Recommendations on the Filtering of IPv6 Packets Containing IPv6 Extension Headers at Transit Routers", Work in Progress, Internet-Draft, draft-ietf-opsec-ipv6-eh-filtering-08, 3 June 2021, <https://datatracker.ietf.org/doc/html/draft-ietf-opsec-ipv6-eh-filtering-08>.

[I-D.draft-vyncke-opsec-probe-attribution] Vyncke, É., Donnet, B., and J. Iurman, "Attribution of Internet Probes", Work in Progress, Internet-Draft, draft-vyncke-opsec-probe-attribution-01, 3 March 2022, <https://datatracker.ietf.org/doc/html/draft-vyncke-opsec-probe-attribution-01>.

[MLAT_PEERING] Giotsas, V., Zhou, S., Luckie, M., and K. Claffy, "Inferring Multilateral Peering", DOI 10.1145/2535372.2535390, December 2013, <https://catalog.caida.org/details/paper/2013_inferring_multilateral_peering/>.

[RFC1753] Chiappa, N., "IPng Technical Requirements Of the Nimrod Routing and Addressing Architecture", RFC 1753, DOI 10.17487/RFC1753, December 1994, <https://www.rfc-editor.org/rfc/rfc1753>.

[RFC4942] Davies, E., Krishnan, S., and P. Savola, "IPv6 Transition/Co-existence Security Considerations", RFC 4942, DOI 10.17487/RFC4942, September 2007, <https://www.rfc-editor.org/rfc/rfc4942>.

[RFC5095] Abley, J., Savola, P., and G. Neville-Neil, "Deprecation of Type 0 Routing Headers in IPv6", RFC 5095, DOI 10.17487/RFC5095, December 2007, <https://www.rfc-editor.org/rfc/rfc5095>.

[RFC6275] Perkins, C., Ed., Johnson, D., and J. Arkko, "Mobility Support in IPv6", RFC 6275, DOI 10.17487/RFC6275, July 2011, <https://www.rfc-editor.org/rfc/rfc6275>.

[RFC6554] Hui, J., Vasseur, JP., Culler, D., and V. Manral, "An IPv6 Routing Header for Source Routes with the Routing Protocol for Low-Power and Lossy Networks (RPL)", RFC

6554, DOI 10.17487/RFC6554, March 2012, <https://www.rfc-editor.org/rfc/rfc6554>.

[RFC7704] Crocker, D. and N. Clark, "An IETF with Much Diversity and Professional Conduct", RFC 7704, DOI 10.17487/RFC7704, November 2015, <https://www.rfc-editor.org/rfc/rfc7704>.

[RFC7872] Gont, F., Linkova, J., Chown, T., and W. Liu, "Observations on the Dropping of Packets with IPv6 Extension Headers in the Real World", RFC 7872, DOI 10.17487/RFC7872, June 2016, <https://www.rfc-editor.org/rfc/rfc7872>.

[RFC8754] Filsfils, C., Ed., Dukes, D., Ed., Previdi, S., Leddy, J., Matsushima, S., and D. Voyer, "IPv6 Segment Routing Header (SRH)", RFC 8754, DOI 10.17487/RFC8754, March 2020, <https://www.rfc-editor.org/rfc/rfc8754>.

[RFC8986] Filsfils, C., Ed., Camarillo, P., Ed., Leddy, J., Voyer, D., Matsushima, S., and Z. Li, "Segment Routing over IPv6 (SRv6) Network Programming", RFC 8986, DOI 10.17487/RFC8986, February 2021, <https://www.rfc-editor.org/rfc/rfc8986>.

[TIER1] "Tier 1 network", n.d., <https://en.wikipedia.org/wiki/Tier_1_network>.

Acknowledgments

The authors want to thank Sander Steffann and Jan Zorz for allowing the free use of their labs. Other thanks to Fernando Gont who indicated a nice IPv6 hosting provider in South America.

Special thanks as well to Professor Benoit Donnet for his support and advices. This document would not have existed without his support.

Authors' Addresses

Éric Vyncke
Cisco
De Kleetlaan 64
1831 Diegem
Belgium

Email: evyncke@cisco.com

Raphaël Léas
Université de Liège

Liège
Belgium

Email: raphael.leas@student.uliege.be

Justin Iurman
Université de Liège
Institut Montefiore B28
Allée de la Découverte 10
4000 Liège
Belgium

Email: justin.iurman@uliege.be