IPv6 Operations Working Group (v6ops)
Internet-Draft
Intended status: Informational
Expires: September 12, 2016

F. Gont
SI6 Networks / UTN-FRH
N. Hilliard
INEX
G. Doering
SpaceNet AG
W. Liu
Huawei Technologies
W. Kumari
Google

March 11, 2016

Operational Implications of IPv6 Packets with Extension Headers
draft-gont-v6ops-ipv6-ehs-packet-drops-03

Abstract

This document summarizes the security and operational implications of IPv6 extension headers, and attempts to analyze reasons why packets with IPv6 extension headers may be dropped in the public Internet.

Table of Contents

1. Introduction 2
2. Previous Work on IPv6 Extension Headers 3
3. Security Implications 4
4. Operational Implications 5
4.1. Requirement to process required layer-3/layer-4 information 5
4.2. Route-Processor Protection 7
4.3. Inability to Perform Fine-grained Filtering 8
5. A Possible Attack Vector 8
6. Future Work 10
7. IANA Considerations 10
8. Security Considerations 10
9. Acknowledgements 11
10. References 11
10.1. Normative References 11
10.2. Informative References 12
Authors' Addresses 15

1. Introduction

IPv6 Extension Headers (EHs) allow for the extension of the IPv6 protocol, and provide support for core functionality such as IPv6 fragmentation. However, common implementation limitations suggest that EHs present a challenge for IPv6 packet routing equipment, and evidence exists to suggest that IPv6 packets with EHs may be intentionally dropped on the public Internet in some network deployments.

The authors of this document have been involved in numerous discussions about IPv6 extension headers (both within the IETF and outside of it), and have noticed that a number of security and operational issues were unknown to the larger audience participating in these discussions.

This document has the following goals:

- Raise awareness about the security and operational implications of IPv6 Extension Headers, and presents reasons why some networks intentionally drop packets containing IPv6 Extension Headers.

- Highlight areas where current IPv6 support by networking devices maybe sub-optimal, such that the aforementioned support is improved.

- Highlight operational issues associated with IPv6 extension headers, such that those issues are considered in IETF standardization efforts.

Section 2 of this document summarizes the previous work that has been done in the area of IPv6 extension headers. Section 3 briefly discusses the security implications of IPv6 Extension Headers, while Section 4 discusses their operational implications. Finally, Section 6 proposes an action plan for improving the state of affairs of IPv6 extension headers.

2. Previous Work on IPv6 Extension Headers

Some of the implications of IPv6 Extension Headers have been discussed in IETF circles. For example, [I-D.taylor-v6ops-fragdrop] discusses a rationale for which operators drop IPv6 fragments. [I-D.wkumari-long-headers] discusses possible issues arising from "long" IPv6 header chains. [RFC7045] clarifies how intermediate nodes should deal with IPv6 extension headers. [RFC7112] discusses the issues arising in a specific fragmentation case where the IPv6 header chain is fragmented into two or more fragments (and formally forbids such fragmentation case). [I-D.kampanakis-6man-ipv6-eh-parsing] describes how inconsistencies in the way IPv6 packets with extension headers are parsed by different implementations may result in evasion of security controls, and presents guidelines for parsing IPv6 extension headers with the goal of providing a common and consistent parsing methodology for IPv6 implementations. [RFC6980] analyzes the security implications of employing IPv6 fragmentation with Neighbor Discovery for IPv6, and formally recommends against such usage. Finally, [RFC7123] discusses how some popular RA-Guard implementations are subject to evasion by means of IPv6 extension headers.

Some preliminary measurements regarding the extent to which packet containing IPv6 EHs are dropped in the public Internet have been presented in [PMTUD-Blackholes], [Gont-IEPG88], [Gont-Chown-IEPG89], and [Linkova-Gont-IEPG90]. [I-D.ietf-v6ops-ipv6-ehs-in-real-world] presents more comprehensive results and documents the methodology for obtaining the presented results.

3. Security Implications

The security implications of IPv6 Extension Headers generally fall into one or more of these categories:

- Evasion of security controls
- DoS due to processing requirements
- DoS due to implementation errors
- Extension Header-specific issues

Unlike IPv4 packets where the upper-layer protocols can be trivially found by means of the "IHL" ("Internet Header Length") IPv4 header field, the structure of IPv6 packets is more flexible and complex. Locating upper-layer protocol information requires that all IPv6 extension headers be examined. This has presented implementation difficulties, and packet filtering mechanisms that require upper-layer information (even if just the upper layer protocol type) on several security devices can be trivially evaded by inserting IPv6 Extension Headers between the main IPv6 header and the upper layer protocol. [RFC7113] describes this issue for the RA-Guard case, but the same techniques can be employed to circumvent other IPv6 firewall and packet filtering mechanisms. Additionally, implementation inconsistencies in packet forwarding engines may result in evasion of security controls [I-D.kampanakis-6man-ipv6-eh-parsing] [Atlasis2014] [BH-EU-2014].

Packets that use IPv6 Extension Headers may have a negative performance impact on the handling devices. Unless appropriate mitigations are put in place (e.g., packet dropping and/or rate-limiting), an attacker could simply send a large amount of IPv6 traffic employing IPv6 Extension Headers with the purpose of performing a Denial of Service (DoS) attack (see Section 4 for further details).

NOTE:
In the most trivial case, a packet that includes a Hop-by-Hop Options header will typically go through the slow forwarding path, and be processed by the router’s CPU. Another possible case might be that in which a router that has been configured to enforce an ACL based on upper-layer information (e.g., upper layer protocol or TCP Destination Port), needs to process the entire IPv6 header chain (in order to find the required information) and this causes the packet to be processed in the slow path [Cisco-EH-Cons]. We note that, for obvious reasons, the aforementioned performance issues may also affect other devices such as firewalls, Network

Intrusion Detection Systems (NIDS), etc. [Zack-FW-Benchmark]. The extent to which these devices are affected will typically be implementation-dependent.

IPv6 implementations, like all other software, tend to mature with time and wide-scale deployment. While the IPv6 protocol itself has existed for almost 20 years, serious bugs related to IPv6 Extension Header processing continue to be discovered. Because there is currently little operational reliance on IPv6 Extension headers, the corresponding code paths are rarely exercised, and there is the potential that bugs still remain to be discovered in some implementations.

IPv6 Fragment Headers are employed to allow fragmentation of IPv6 packets. While many of the security implications of the fragmentation / reassembly mechanism are known from the IPv4 world, several related issues have crept into IPv6 implementations. These range from denial of service attacks to information leakage, for example [I-D.ietf-6man-predictable-fragment-id], [Bonica-NANOG58] and [Atlasis2012]).

4. Operational Implications

4.1. Requirement to process required layer-3/layer-4 information

Intermediate systems and middleboxes that need to find the layer-4 header must process the entire IPv6 extension header chain. When such devices are unable to obtain the required information, they may simply drop the corresponding packets. The following subsections discuss some of reasons for which such layer-4 information may be needed by an intermediate systems or middlebox, and why packets containing IPv6 extension headers may represent a challenge in such scenarios.

4.1.1. Packet Forwarding Engine Constraints

Most modern routers use dedicated hardware (e.g. ASICs or NPUs) to determine how to forward packets across their internal fabrics (see [IEPG94-Scudder] for details). One of the common methods of handling next-hop lookup is to send a small portion of the ingress packet to a lookup engine with specialised hardware (e.g. ternary CAM or RLDRAM) to determine the packet’s next-hop. Technical constraints mean that there is a trade-off between the amount of data sent to the lookup engine and the overall performance of the lookup engine. If more data is sent, the lookup engine can inspect further into the packet, but the overall performance of the system will be reduced. If less data is sent, the overall performance of the router will be increased

but the packet lookup engine may not be able to inspect far enough into a packet to determine how it should be handled.

NOTE:
For example, current high-end routers at the time of authorship of this document can use up to 192 bytes of header (Cisco ASR9000 Typhoon) or 384 bytes of header (Juniper MX Trio)

If a hardware forwarding engine on a modern router cannot make a forwarding decision about a packet because critical information is not sent to the look-up engine, then the router will normally drop the packet. Historically, some packet forwarding engines punted packets of this form to the control plane for more in-depth analysis, but this is unfeasible on most current router architectures as a result of the vast difference between the hardware forwarding capacity of the router and processing capacity of the control plane and the size of the management link which connects the control plane to the forwarding plane.

If an IPv6 header chain is sufficiently long that its header exceeds the packet look-up capacity of the router, then it may be dropped due to hardware inability to determine how it should be handled.

4.1.2. ECMP and Hash-based Load-Sharing

In the case of ECMP (equal cost multi path) load sharing, the router on the sending side of the link needs to make a decision regarding which of the links to use for a given packet. Since round-robin usage of the links is usually avoided in order to prevent packet reordering, forwarding engines need to use a mechanism which will consistently forward the same data streams down the same forwarding paths. Most forwarding engines achieve this by calculating a simple hash using an n-tuple gleaned from a combination of layer-2 through to layer-4 packet header information. This n-tuple will typically use the src/dst MAC address, src/dst IP address, and if possible further layer-4 src/dst port information. As layer-4 port information increases the entropy of the hash, it is highly desirable to use it where possible.

We note that in the IPv6 world, flows are expected to be identified by means of the IPv6 Flow Label [RFC6437]. Thus, ECMP and Hash-based Load-Sharing would be possible without the need to process the entire IPv6 header chain to obtain upper-layer information to identify flows. However, we note that for a long time many IPv6 implementations failed to set the Flow Label, and ECMP and Hash-based Load-Sharing devices also did not employ the Flow Label for performing their task.

Clearly, widespread support of [RFC6437] would relieve middle-boxes from having to process the entire IPv6 header chain, making Flow Label-based ECMP and Hash-based Load-Sharing [RFC6438] feasible.

4.1.3. Enforcing infrastructure ACLs

Generally speaking, infrastructure ACLs (iACLs) drop unwanted packets destined to parts of a provider's infrastructure, because they are not operationally needed and can be used for attacks of different sorts against the router's control plane. Some traffic needs to be differentiated depending on layer-3 or layer-4 criteria to achieve a useful balance of protection and functionality, for example:

- Permit some amount of ICMP echo (ping) traffic towards the router's addresses for troubleshooting.

- Permit BGP sessions on the shared network of an exchange point (potentially differentiating between the amount of packets/seconds permitted for established sessions and connection establishment), but do not permit other traffic from the same peer IP addresses.

4.1.4. DDoS Management and Customer Requests for Filtering

The case of customer DDoS protection and edge-to-core customer protection filters is similar in nature to the infrastructure ACL protection. Similar to infrastructure ACL protection, layer-4 ACLs generally need to be applied as close to the edge of the network as possible, even though the intent is usually to protect the customer edge rather than the provider core. Application of layer-4 DDoS protection to a network edge is often automated using Flowspec [RFC5575].

For example, a web site which normally only handled traffic on TCP ports 80 and 443 could be subject to a volumetric DDoS attack using NTP and DNS packets with randomised source IP address, thereby rendering useless traditional [RFC5635] source-based real-time black hole mechanisms. In this situation, DDoS protection ACLs could be configured to block all UDP traffic at the network edge without impairing the web server functionality in any way. Thus, being able to block arbitrary protocols at the network edge can avoid DDoS-related problems both in the provider network and on the customer edge link.

4.2. Route-Processor Protection

Most modern routers have a fast hardware-assisted forwarding plane and a loosely coupled control plane, connected together with a link that has much less capacity than the forwarding plane could handle.

Traffic differentiation cannot be done by the control plane side, because this would overload the internal link connecting the forwarding plane to the control plane.

The Hop-by-Hop Options header is particularly challenging since, in most (if not all) implementations, it causes the corresponding packet to be punted to a software path. As a result, operators usually drop IPv6 packets containing this extension header. Please see [RFC6192] for advice regarding protection of the router control plane.

4.3. Inability to Perform Fine-grained Filtering

Some routers lack of fine-grained filtering of IPv6 extension headers. For example, an operator may want to drop packets containing Routing Header Type 0 (RHT0) but may only be able to filter on the extension header type (Routing Header). As a result, the operator may end up enforcing a more coarse filtering policy (e.g. "drop all packets containing a Routing Header" vs. "only drop packets that contain a Routing Header Type 0").

5. A Possible Attack Vector

The widespread drop of IPv6 packets employing IPv6 Extension Headers can, in some scenarios, be exploited for malicious purposes: if packets employing IPv6 EHs are known to be dropped on the path from system A to system B, an attacker could cause packets sent from A to B to be dropped by sending a forged ICMPv6 Packet Too Big (PTB) [RFC4443] error message to A (advertising an MTU smaller than 1280), such that subsequent packets from A to B include a fragment header (i.e., they result in atomic fragments [RFC6946]).

Possible scenarios where this attack vector could be exploited include (but are not limited to):

- Communication between any two systems through the public network (e.g., client from/to server or server from/to server), where packets with IPv6 extension headers are dropped by some intermediate router

- Communication between two BGP peers employing IPv6 transport, where these BGP peers implement ACLs to drop IPv6 fragments (to avoid control-plane attacks)

The aforementioned attack vector is exacerbated by the following factors:

- The attacker does not need to forge the IPv6 Source Address of his attack packets. Hence, deployment of simple BCP38 filters will not help as a counter-measure.

- Only the IPv6 addresses of the IPv6 packet embedded in the ICMPv6 payload need to be forged. While one could envision filtering devices enforcing BCP38-style filters on the ICMPv6 payload, the use of extension headers (by the attacker) could make this difficult, if not impossible.

- Many implementations fail to perform validation checks on the received ICMPv6 error messages, as recommended in Section 5.2 of [RFC4443] and documented in [RFC5927]. It should be noted that in some cases, such as when an ICMPv6 error message has (supposedly) been elicited by a connection-less transport protocol (or some other connection-less protocol being encapsulated in IPv6), it may be virtually impossible to perform validation checks on the received ICMPv6 error messages. And, because of IPv6 extension headers, the ICMPv6 payload might not even contain any useful information on which to perform validation checks.

- Upon receipt of one of the aforementioned ICMPv6 "Packet Too Big" error messages, the Destination Cache [RFC4861] is usually updated to reflect that any subsequent packets to such destination should include a Fragment Header. This means that a single ICMPv6 "Packet Too Big" error message might affect multiple communication instances (e.g. TCP connections) with such destination.

- A router or other middlebox cannot simply drop all incoming ICMPv6 Packet Too Big error messages, as this would create a PMTUD blackhole.

Possible mitigations for this issue include:

- Dropping incoming ICMPv6 Packet Too Big error messages that advertise an MTU smaller than 1280 bytes.

- Artificially reducing the MTU to 1280 bytes and dropping incoming ICMPv6 PTB error messages.

Both of these mitigations come at the expense of possibly preventing communication through SIIT [RFC6145], that relies on IPv6 atomic fragments (see [I-D.ietf-6man-deprecate-atomfrag-generation]), and also implies that the filtering device has the ability to filter ICMP PTB messages based on the contents of the MTU field.

[I-D.ietf-6man-deprecate-atomfrag-generation] documents while the generation of IPv6 atomic fragments is considered harmful, and

documents why this functionality is being removed from the upcoming revision of the core IPv6 protocol [I-D.ietf-6man-rfc2460bis]. Thus, any of the above mitigations would eliminate the attack vector without any interoperability implications.

6. Future Work

Based on the discussion provided in this document, we recommend the following (*non*-mutually exclusive) actions to improve the state of affairs of IPv6 extension headers:

- Vendors must allow for better granularity in the specification of filters for IPv6 extension headers, such that filters for specific EH types and subtypes (e.g. RHT0 vs. RHT2) can be specified wthout affecting other extension header types/subtypes unnecessarily (please see Section 4.3).

- Provide advice on the filtering of IPv6 packets that contain IPv6 extension headers (as in [I-D.ietf-opsec-ipv6-eh-filtering]).

- The IETF should evaluate the possibility of enforcing a cap on the maximum length of an IPv6 EH chain (e.g., as proposed in [I-D.wkumari-long-headers]). If not at the protocol specification level (i.e., "Standards Track"), such a cap could be recommended as operational advice of the form "IPv6 implementations are expected to support EH chains as long as the they fit in the Path-MTU for the corresponding packets (see [RFC7112]). However, given current technology constraints, we specifically note that all implementations MUST support EH chains of at least X bytes, and MUST be able to process such EH chains (where necessary), without negative performance impact".

We explicitly note that the authors of this document do not (in any way) suggest or propose to depracate IPv6 extension headers and that, on the contrary, they propose actions to improve their state of affairs.

7. IANA Considerations

There are no IANA registries within this document. The RFC-Editor can remove this section before publication of this document as an RFC.

8. Security Considerations

The security implications of IPv6 extension headers are discussed in Section 3. A specific attack vector that could leverage the widespread dropping of packets with IPv6 EHs (along with possible

countermeasures) is discussed in Section 5. This document does not introduce any new security issues.

9. Acknowledgements

The authors would like to thank (in alphabetical order) Mikael Abrahamsson, Fred Baker, Brian Carpenter, Lee Howard, Sander Steffann, Eric Vyncke, and Andrew Yourtchenko, for providing valuable comments on earlier versions of this document. Additionally, the authors would like to thank participants of the v6ops working group for their valuable input on the topics that led to the publication of this document.

Fernando Gont would like to thank Sander Steffann, who took the time to meet to discuss this document, even while higher priority events were in place.

Fernando Gont would like to thank Jan Zorz / Go6 Lab <http://go6lab.si/>, and Jared Mauch / NTT America, for providing access to systems and networks that were employed to perform experiments and measurements involving packets with IPv6 Extension Headers. Additionally, he would like to thank SixXS <https://www.sixxs.net> for providing IPv6 connectivity.

10. References

10.1. Normative References

[RFC2460] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", RFC 2460, DOI 10.17487/RFC2460, December 1998, <http://www.rfc-editor.org/info/rfc2460>.

[RFC4443] Conta, A., Deering, S., and M. Gupta, Ed., "Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification", RFC 4443, DOI 10.17487/RFC4443, March 2006, <http://www.rfc-editor.org/info/rfc4443>.

[RFC4861] Narten, T., Nordmark, E., Simpson, W., and H. Soliman, "Neighbor Discovery for IP version 6 (IPv6)", RFC 4861, DOI 10.17487/RFC4861, September 2007, <http://www.rfc-editor.org/info/rfc4861>.

[RFC6145] Li, X., Bao, C., and F. Baker, "IP/ICMP Translation Algorithm", RFC 6145, DOI 10.17487/RFC6145, April 2011, <http://www.rfc-editor.org/info/rfc6145>.

[RFC6946] Gont, F., "Processing of IPv6 "Atomic" Fragments", RFC 6946, DOI 10.17487/RFC6946, May 2013, <http://www.rfc-editor.org/info/rfc6946>.

10.2. Informative References

[Atlasis2012]
Atlasis, A., "Attacking IPv6 Implementation Using Fragmentation", BlackHat Europe 2012. Amsterdam, Netherlands. March 14-16, 2012, <https://media.blackhat.com/bh-eu-12/Atlasis/bh-eu-12-Atlasis-Attacking_IPv6-Slides.pdf>.

[Atlasis2014]
Atlasis, A., "A Novel Way of Abusing IPv6 Extension Headers to Evade IPv6 Security Devices", May 2014, <http://www.insinuator.net/2014/05/a-novel-way-of-abusing-ipv6-extension-headers-to-evade-ipv6-security-devices/>.

[BH-EU-2014]
Atlasis, A., Rey, E., and R. Schaefer, "Evasion of High-End IDPS Devices at the IPv6 Era", BlackHat Europe 2014, 2014, <https://www.ernw.de/download/eu-14-Atlasis-Rey-Schaefer-briefings-Evasion-of-HighEnd-IPS-Devices-wp.pdf>.

[Bonica-NANOG58]
Bonica, R., "IPv6 Extension Headers in the Real World v2.0", NANOG 58. New Orleans, Louisiana, USA. June 3-5, 2013, <https://www.nanog.org/sites/default/files/mon.general.fragmentation.bonica.pdf>.

[Cisco-EH-Cons]
Cisco, , "IPv6 Extension Headers Review and Considerations", October 2006, <http://www.cisco.com/en/US/technologies/tk648/tk872/technologies_white_paper0900aecd8054d37d.pdf>.

[Gont-Chown-IEPG89]
Gont, F. and T. Chown, "A Small Update on the Use of IPv6 Extension Headers", IEPG 89. London, UK. March 2, 2014, <http://www.iepg.org/2014-03-02-ietf89/fgont-iepg-ietf89-eh-update.pdf>.

[Gont-IEPG88]
Gont, F., "Fragmentation and Extension header Support in the IPv6 Internet", IEPG 88. Vancouver, BC, Canada. November 13, 2013, <http://www.iepg.org/2013-11-ietf88/fgont-iepg-ietf88-ipv6-frag-and-eh.pdf>.

[I-D.ietf-6man-deprecate-atomfrag-generation]
Gont, F., LIU, S., and T. Anderson, "Generation of IPv6 Atomic Fragments Considered Harmful", draft-ietf-6man-deprecate-atomfrag-generation-05 (work in progress), January 2016.

[I-D.ietf-6man-predictable-fragment-id]
Gont, F., "Security Implications of Predictable Fragment Identification Values", draft-ietf-6man-predictable-fragment-id-10 (work in progress), October 2015.

[I-D.ietf-6man-rfc2460bis]
Deering, S. and B. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", draft-ietf-6man-rfc2460bis-03 (work in progress), January 2016.

[I-D.ietf-opsec-ipv6-eh-filtering]
Gont, F., LIU, S., and R. Bonica, "Recommendations on Filtering of IPv6 Packets Containing IPv6 Extension Headers", draft-ietf-opsec-ipv6-eh-filtering-00 (work in progress), March 2015.

[I-D.ietf-v6ops-ipv6-ehs-in-real-world]
Gont, F., Linkova, J., Chown, T., and S. LIU, "Observations on the Dropping of Packets with IPv6 Extension Headers in the Real World", draft-ietf-v6ops-ipv6-ehs-in-real-world-02 (work in progress), December 2015.

[I-D.kampanakis-6man-ipv6-eh-parsing]
Kampanakis, P., "Implementation Guidelines for parsing IPv6 Extension Headers", draft-kampanakis-6man-ipv6-eh-parsing-01 (work in progress), August 2014.

[I-D.taylor-v6ops-fragdrop]
Jaeggli, J., Colitti, L., Kumari, W., Vyncke, E., Kaeo, M., and T. Taylor, "Why Operators Filter Fragments and What It Implies", draft-taylor-v6ops-fragdrop-02 (work in progress), December 2013.

[I-D.wkumari-long-headers]
Kumari, W., Jaeggli, J., Bonica, R., and J. Linkova, "Operational Issues Associated With Long IPv6 Header Chains", draft-wkumari-long-headers-03 (work in progress), June 2015.

[IEPG94-Scudder]
Petersen, B. and J. Scudder, "Modern Router Architecture for Protocol Designers", IEPG 94. Yokohama, Japan. November 1, 2015, <http://www.iepg.org/2015-11-01-ietf94/IEPG-RouterArchitecture-jgs.pdf>.

[Linkova-Gont-IEPG90]
Linkova, J. and F. Gont, "IPv6 Extension Headers in the Real World v2.0", IEPG 90. Toronto, ON, Canada. July 20, 2014, <http://www.iepg.org/2014-07-20-ietf90/iepg-ietf90-ipv6-ehs-in-the-real-world-v2.0.pdf>.

[PMTUD-Blackholes]
De Boer, M. and J. Bosma, "Discovering Path MTU black holes on the Internet using RIPE Atlas", July 2012, <http://www.nlnetlabs.nl/downloads/publications/pmtu-black-holes-msc-thesis.pdf>.

[RFC5575] Marques, P., Sheth, N., Raszuk, R., Greene, B., Mauch, J., and D. McPherson, "Dissemination of Flow Specification Rules", RFC 5575, DOI 10.17487/RFC5575, August 2009, <http://www.rfc-editor.org/info/rfc5575>.

[RFC5635] Kumari, W. and D. McPherson, "Remote Triggered Black Hole Filtering with Unicast Reverse Path Forwarding (uRPF)", RFC 5635, DOI 10.17487/RFC5635, August 2009, <http://www.rfc-editor.org/info/rfc5635>.

[RFC5927] Gont, F., "ICMP Attacks against TCP", RFC 5927, DOI 10.17487/RFC5927, July 2010, <http://www.rfc-editor.org/info/rfc5927>.

[RFC6192] Dugal, D., Pignataro, C., and R. Dunn, "Protecting the Router Control Plane", RFC 6192, DOI 10.17487/RFC6192, March 2011, <http://www.rfc-editor.org/info/rfc6192>.

[RFC6437] Amante, S., Carpenter, B., Jiang, S., and J. Rajahalme, "IPv6 Flow Label Specification", RFC 6437, DOI 10.17487/RFC6437, November 2011, <http://www.rfc-editor.org/info/rfc6437>.

[RFC6438] Carpenter, B. and S. Amante, "Using the IPv6 Flow Label for Equal Cost Multipath Routing and Link Aggregation in Tunnels", RFC 6438, DOI 10.17487/RFC6438, November 2011, <http://www.rfc-editor.org/info/rfc6438>.

[RFC6980] Gont, F., "Security Implications of IPv6 Fragmentation with IPv6 Neighbor Discovery", RFC 6980, DOI 10.17487/RFC6980, August 2013, <http://www.rfc-editor.org/info/rfc6980>.

[RFC7045] Carpenter, B. and S. Jiang, "Transmission and Processing of IPv6 Extension Headers", RFC 7045, DOI 10.17487/RFC7045, December 2013, <http://www.rfc-editor.org/info/rfc7045>.

[RFC7112] Gont, F., Manral, V., and R. Bonica, "Implications of Oversized IPv6 Header Chains", RFC 7112, DOI 10.17487/RFC7112, January 2014, <http://www.rfc-editor.org/info/rfc7112>.

[RFC7113] Gont, F., "Implementation Advice for IPv6 Router Advertisement Guard (RA-Guard)", RFC 7113, DOI 10.17487/RFC7113, February 2014, <http://www.rfc-editor.org/info/rfc7113>.

[RFC7123] Gont, F. and W. Liu, "Security Implications of IPv6 on IPv4 Networks", RFC 7123, DOI 10.17487/RFC7123, February 2014, <http://www.rfc-editor.org/info/rfc7123>.

[RIPE-Atlas] RIPE, , "RIPE Atlas", <https://atlas.ripe.net/>.

[Zack-FW-Benchmark] Zack, E., "Firewall Security Assessment and Benchmarking IPv6 Firewall Load Tests", IPv6 Hackers Meeting #1, Berlin, Germany. June 30, 2013, <http://www.ipv6hackers.org/meetings/ipv6-hackers-1/zack-ipv6hackers1-firewall-security-assessment-and-benchmarking.pdf>.

Authors' Addresses

Fernando Gont
SI6 Networks / UTN-FRH
Evaristo Carriego 2644
Haedo, Provincia de Buenos Aires 1706
Argentina

Phone: +54 11 4650 8472
Email: fgont@si6networks.com
URI: http://www.si6networks.com

Nick Hilliard
INEX
4027 Kingswood Road
Dublin 24
IE

Email: nick@inex.ie

Gert Doering
SpaceNet AG
Joseph-Dollinger-Bogen 14
Muenchen D-80807
Germany

Email: gert@space.net

Will (Shucheng) Liu
Huawei Technologies
Bantian, Longgang District
Shenzhen 518129
P.R. China

Email: liushucheng@huawei.com

Warren Kumari
Google
1600 Amphitheatre Parkway
Mountain View, CA 94043
US

Email: warren@kumari.net

Internet Engineering Task Force
Internet-Draft
Intended status: Informational
Expires: August 30, 2016

M. Smith

February 27, 2016

Further Mitigating Router ND Cache Exhaustion DoS Attacks Using Solicited-Node Group Membership
draft-smith-v6ops-mitigate-rtr-dos-mld-slctd-node-02

Abstract

For each of their IPv6 unicast or anycast addresses, nodes join a Solicited-Node multicast group, formed using the lower 24 bits of the address. This Solicited-Node group membership could be used by routers to further mitigate a Neighbor Discovery cache Denial of Service attack.

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 http://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 August 30, 2016.

Copyright Notice

Table of Contents

1. Introduction 2
2. Method 4
2.1. Tracking Solicited-Node Multicast Group Presence 4
2.2. Neighbor Presence Discovery 5
2.2.1. Strict Mitigation Mode 5
2.2.2. Relaxed Mitigation Mode 6
3. MLD Reliability 6
4. Security Considerations 8
5. Acknowledgements 9
6. Change Log [RFC Editor please remove] 9
7. Informative References 10
Author's Address 11

1. Introduction

When an IPv6 unicast or anycast address is added to or removed from an interface, a node is also required to join or leave the Solicited-Node multicast group that corresponds to the address [RFC4291][RFC6434], using the Multicast Listener Discovery (MLD) protocol [RFC2710][RFC3810]. The Solicited-Node multicast group the node joins or leaves is determined by appending the lower 24 bits of the unicast or anycast address, usually part of the Interface Identifier (IID), to the IPv6 multicast prefix FF02:0:0:0:0:1:FF00::/104 [RFC4291].

The current use of Solicited-Node multicast groups is to avoid having to link layer broadcast Neighbor Discovery (ND) Neighbor Solicitations to all nodes on the link (ARP's [RFC0826] behaviour for most ARP Requests). Instead, Neighbor Solicitations are sent to the Solicited-Node multicast group formed from the target address of the Neighbor Solicitation.

The use of Solicited-Node multicast groups for Neighbor Solicitations allows nodes to possibly filter Neighbor Solicitations they aren't interested in in their link layer network interface, avoiding interrupting the node's general purpose CPU (see sections 7.4 and 7.5 of [RFC1112] for further discussion), and possibly for the link layer forwarding device(s) to avoid sending Neighbor Solicitations to nodes that do not have the target address [RFC4541]. Facilitating link layer network interface multicast filtering and reducing the flooding scope of multicasts on a link helps increase the number of nodes that can be attached to a link.

As the addition or removal of unicast or anycast addresses triggers Solicited-Node multicast group joins or leaves, this mechanism is in effect a low resolution address range presence registration protocol, registering portions of the on-link address range for which there are unicast or anycast addresses present. The presence of a Solicited-Node multicast group on a link indicates that at least one unicast or anycast address that maps to the Solicited-Node multicast group is present. Conversely, the absence of a Solicited-Node multicast group on a link indicates that no unicast or anycast addresses are present that would map to the corresponding Solicited-Node multicast group.

MLDv2 joins for Solicited-Node multicast groups could also be used as a link-local address registration method for at least one of each nodes' link-local addresses, as link-local unicast addresses are used as MLDv2 source addresses, excepting MLDv2 joins for Solicited-Node multicast groups when a link-local address is not available [RFC3590]. It would not be possible to do this reliably with MLDv1 Solicited-Node multicast group joins as MLDv1 listeners will suppress joins for their own groups if they hear a join for the same groups from another listener.

This presence or absence of Solicited-Node multicast groups could be used by a router to determine if it needs to send Neighbor Solicitations for unresolved addresses on to the link. If the to-be-resolved address maps to a non-existent Solicited-Node multicast group, the router could discard the packet, rather than sending a Neighbor Solicitation to the corresponding Solicited-Node multicast group for the packet's destination and possibly queuing the trigger packet while neighbor discovery occurs. Discarding trigger packets that map to absent Solicited-Node multicast groups could be a further Neighbor Discovery cache Denial of Service (DoS) attack [RFC3756] mitigation technique.

For links with prefixes with lengths shorter than or equal to /104, such as the common /64 [RFC7421], the total number of Solicited-Node multicast groups possible on a link is 2^24, or 16 777 216 groups. The number of Solicited-Node multicast groups present on a link is equal to the number of IPv6 unicast or anycast addresses present on the link which have unique lower 24 bits, used to form the Solicited-Node multicast group address.

For most links the number of present Solicited-Node multicast groups present will be in the order of 10s, 100s or perhaps on rarer occasions in the low 1000s. This means that Neighbor Solicitations do not have to be sent for very large numbers of unresolved unicast or anycast addresses for which the corresponding Solicited-Node multicast group is not present. This would significantly reduce the attack surface for the ND cache exhaustion denial of service attack.

For example, if a link has 1000 present Solicited-Node multicast groups, then Neighbor Solicitations do not have to be sent for addresses that would map to the absent 16 776 216 Solicited-Node multicast groups, more than 99.99% of the possible on-link Solicited-Node multicast groups.

This memo describes how a router could collect Solicited-Node multicast group membership and how it could use this information as part of its neighbor presence discovery procedure, for the purposes of further mitigating the ND cache exhaustion attack.

Note that this method has been independently suggested by Greg Daley and perhaps others.

2. Method

2.1. Tracking Solicited-Node Multicast Group Presence

To track Solicited-Node multicast group presence on a link, a router uses the multicast listener discovery procedures specified in [RFC2710] or [RFC3810], without modification.

Note that the procedures specified in [RFC2710] and [RFC3810] do not require that a router performing them is to forward multicast packets, or is to be participating in a multicast routing protocol with other multicast routers. The ND cache DoS mitigation method described in this memo can be used regardless of whether the other routers in the network, including other on-link routers, are performing multicast forwarding.

If a router using this ND cache DoS mitigation method is not performing multicast forwarding, it may choose to only track Solicited-Node multicast group presence, ignoring the presence information it receives for other multicast groups. This may usefully reduce the router’s resources consumption. If a router using this optimisation becomes a multicast forwarding router, it will need to collect presence information for all on-link multicast groups, using the Querier Election procedure [RFC2710][RFC3810], as though it had just been attached to the link, and had no knowledge of the presence any of the multicast groups.

A router with two or more interfaces attached to the same link only needs to operate MLD on one of those interfaces [RFC3810]; the list of on-link Solicited-Node multicast groups would be used across all of these interfaces when mitigating ND cache DoSes.

2.2. Neighbor Presence Discovery

When a router receives a packet for a destination for which it does not have a neighbor cache entry, it uses the [RFC4291] specified method to form a Solicited-Node multicast group address from the packet's destination address.

The router then compares the resulting Solicited-Node multicast group address with its list of present Solicited-Node multicast groups on the link.

If the Solicited-Node multicast group is present, the router then performs the address resolution procedure for the packet's destination IPv6 address as specified in [RFC4861], starting with sending a Neighbor Solicitation towards the Solicited-Node multicast group that corresponds to the address.

Alternatively, when the Solicited-Node multicast group is not present, the router operates in one of two mitigation modes.

2.2.1. Strict Mitigation Mode

When operating in Strict Mitigation Mode, the router discards all packets whose destination address Solicited-Node multicast groups do not match any of the Solicited-Node multicast groups present on the link.

Strict Mitigation Mode makes the decision to perform Neighbor Discovery dependent on the successful discovery of the Solicited-Node multicast groups on the link by MLD. This means that if the router is assembling a list of present Solicited-Node multicast groups from scratch, such as after the router has been intialised, or when an interface comes online, there will be a period where Neighbor Discovery for existing nodes will not occur, while the full set of present Solicited-Node multicast groups are discovered. To off-link hosts sending traffic to the possible on-link hosts, this will appear to be a period of packet loss. These hosts are expected to have implemented methods to recover from transient failures of transmission, such as packet retransmission, if necessary [RFC1958].

This mode of operation is appropriate when it is known that all attached nodes announce their Solicited-Node multicast group membership for their addresses, and MLD operation on the link is known to be reliable. An example scenario would be a large Internet content provider's environment, where the content network routers and content servers are being operated by the same organisation.

2.2.2. Relaxed Mitigation Mode

When operating in Relaxed Mitigation Mode, under normal non-DoS circumstances the router will also perform the address resolution procedure for packets whose destination address Solicited-Node multicast group does not match any of the Solicited-Node multicast groups present on the link.

However, when there is an indication that a neighbor cache Denial of Service attack might be occurring, the router treats packets whose destination address Solicited-Node multicast group does not match a link present Solicited-Node multicast group with lower importance to those packets whose do.

Indicators that a neighbor cache Denial of Service attack might be occurring could be many failed address resolution attempts over a short period of time, rapid and unexpected consumption of neighbor cache resources (rapid consumption for a short period of time after the link or router has come on-line could be expected), or some other pattern of neighbor cache Denial of Service attack specific behaviour.

If a neighbor cache Denial of Service attack appears to be occurring, an implementation could immediately start discarding packets whose destination address Solicited-Node multicast group does not match those present on the link. A less harsh alternative would be to start discarding some of these packets, increasing the discard rate as neighbor cache resources are increasingly consumed by the Denial of Service attack.

This mode of operation would be appropriate when it is not known if all nodes will announce their Solicited Node multicast group membership, possibly due to some nodes being pre-[RFC2710] implementations or if MLD operation is not known to be reliable. Example scenarios would be residential or public Internet access networks, where the support for or reliability of MLD joins for Solicited Node multicast groups is not known. Specific to the residential network case, where the technical ability of the router operator is not known and likely to be low, Relaxed Mitigation Mode would be the safest default.

3. MLD Reliability

MLD is currently being used for two purposes:

- to join and leave multicast groups so that multicast applications will receive routed multicast traffic they are interested in receiving [RFC2710][RFC3810], and

- o to advise link layer devices of node multicast group membership to allow the link layer devices to limit to which devices multicast traffic is sent, instead of flooding multicast traffic to all attached devices [RFC4541]. Specific to this memo's topic, nodes using MLD to join Solicited-Node multicast groups for their addresses allows link layer devices to limit to which nodes multicast Neighbor Solicitations are sent.

For the first purpose, partial or complete failure of MLD to successfully join the intended multicast group(s) will likely cause the respective multicast application(s) to not function adequately or completely. While likely to be unacceptable to the application(s') user(s), the effects of the failure are limited to the impacted application(s); some multicast applications may function, and other unicast-only applications will not be impacted.

For the second purpose, partial or complete failure of MLD operation means the link layer device will not forward multicast traffic to devices for groups for which MLD joins failed. As with the first MLD purpose, application operation is likely to be impacted. MLD join failures for Solicited-Node multicast groups would mean that Duplicate Address Detection [RFC4861] and Neighbor Discovery [RFC4861] for the node's addresses will fail. IPv6 unicast connectivity for the effected node could be severely impacted, and possibly fail completely.

For this memo's method, when operating in Strict Mitigation Mode, partial or complete failure of MLD for Solicited-Node multicast group joins will cause Neighbor Discovery to fail for routers implementing this neighbor cache Denial of Service attack mitigation. The effected nodes will be unreachable for traffic sources beyond the impacted router.

With this memo's method, when operating in Relaxed Mitigation Mode, partial or complete failure of MLD for Solicited-Node multicast group joins will cause the router to consider neighbor discovery for the effected node's addresses to be of lower importance. Under normal, non-neighbor cache Denial of Service circumstances, these nodes will receive equal service to those who've successfully joined the Solicited-Node multicast groups via MLD. If a neighbor cache Denial of Service occurs, these MLD failed nodes will either have less success at or complete failure of being discovered by the router performing neighbor discovery. In this situation, some rather than all of the nodes will have been impacted by the Denial of Service attack, which is an improvement over the attack impacting all nodes.

It is important to note that failure of neighbor discovery during a neighbor cache Denial of Service attack will only impact nodes that

have not been previously discovered by the router. If a node has been previously discovered, its neighbor information will already reside in the router's neighbor cache, and its currency will be maintained by Neighbor Unreachability Detection [RFC4861].

Due to the number of significant consequences of MLD failure, including those introduced by this memo's method, MLD should be configured to operate reliably if the default MLD reliability related parameter values are not adequate [RFC2710][RFC3810]. Although [RFC6636] provides advice for tuning MLD operation for mobile and wireless networks, some of the advice and considerations might be more generally applied.

4. Security Considerations

The method described in this memo further mitigates the ND cache exhaustion DoS attack. It does not prevent it.

Using this method, neighbor presence discovery will occur for any of the unicast or anycast addresses that map to the present Solicited-Node multicast groups. As a Solicited-Node multicast group can map to up to 2^40 unicast or anycast addresses (for a /64 prefix, 2^(64 - 24)), the ND implementation is likely to continue to be vulnerable to a ND cache exhaustion denial of service for addresses covered by the present Solicited-Node multicast groups. While the number of non-existent addresses that can be targetted remains very large, it is very significantly smaller than the targettable non-existent addresses possible in the on-link prefixes without this measure.

The severity of this threat depends on two factors:

- the number of Solicited-Node multicast groups present on the link, and
- the ability of the off-link attacker to stumble upon or discover non-existent addresses that map to present Solicited-Node multicast groups.

The severity of the threat is lower with lesser numbers of Solicited-Node multicast groups, and less predictable and sparsely distributed Solicited-Node multicast group addresses.

[RFC7217] specifies the use of stable yet random and unpredictable IIDs, on a per-prefix basis. This will increase the number of present Solicited-Node multicast groups, by up to the number of prefixes multiplied by the number of hosts implementing [RFC7217]. This will reduce the effectiveness of the measure proposed in this memo. However, it will also conversely increase the effectiveness of

this measure, as the IIDs and therefore the Solicited-Node multicast groups become less predictable and more sparsely distributed.

To protect against ND cache DoS attacks for non-existent addresses that map to present Solicited-Node multicast groups, other ND cache protection measures, such as those described in [RFC6583] should also be implemented.

When a packet is sent to a destination that is unresolved and is not covered by a present Solicited-Node multicast group, a copy could be sent to an [RFC6018] greynet collector for further analysis. For example, packet sent to destinations falling outside the present Solicited-Node multicast groups could be an indication of an attempt to discover nodes via address probing.

5. Acknowledgements

Review and comments were provided by (in alphabetical order) Fred Baker, Lorenzo Colitti and Ray Hunter. Lorenzo expressed a concern about MLD reliability and its consequences, which prompted the creation of the two modes of mitigation and the MLD reliability discussion.

This memo was prepared using the xml2rfc tool.

6. Change Log [RFC Editor please remove]

draft-smith-v6ops-mitigate-rtr-dos-mld-slctd-node-00, initial version, 2014-04-08

draft-smith-v6ops-mitigate-rtr-dos-mld-slctd-node-01, 2016-02-08

- two modes of response to possible neighbor cache DoS
- MLD reliability discussion
- Expansion on the purpose of using solicited-node multicast groups

draft-smith-v6ops-mitigate-rtr-dos-mld-slctd-node-02, 2016-02-28

- "discard" rather than "drop" to show intentional dropping
- text about Strict Mode MLD discovery period

7. Informative References

[RFC0826] Plummer, D., "Ethernet Address Resolution Protocol: Or Converting Network Protocol Addresses to 48.bit Ethernet Address for Transmission on Ethernet Hardware", STD 37, RFC 826, DOI 10.17487/RFC0826, November 1982, <http://www.rfc-editor.org/info/rfc826>.

[RFC1112] Deering, S., "Host extensions for IP multicasting", STD 5, RFC 1112, DOI 10.17487/RFC1112, August 1989, <http://www.rfc-editor.org/info/rfc1112>.

[RFC1958] Carpenter, B., Ed., "Architectural Principles of the Internet", RFC 1958, DOI 10.17487/RFC1958, June 1996, <http://www.rfc-editor.org/info/rfc1958>.

[RFC2710] Deering, S., Fenner, W., and B. Haberman, "Multicast Listener Discovery (MLD) for IPv6", RFC 2710, DOI 10.17487/RFC2710, October 1999, <http://www.rfc-editor.org/info/rfc2710>.

[RFC3590] Haberman, B., "Source Address Selection for the Multicast Listener Discovery (MLD) Protocol", RFC 3590, DOI 10.17487/RFC3590, September 2003, <http://www.rfc-editor.org/info/rfc3590>.

[RFC3756] Nikander, P., Ed., Kempf, J., and E. Nordmark, "IPv6 Neighbor Discovery (ND) Trust Models and Threats", RFC 3756, DOI 10.17487/RFC3756, May 2004, <http://www.rfc-editor.org/info/rfc3756>.

[RFC3810] Vida, R., Ed. and L. Costa, Ed., "Multicast Listener Discovery Version 2 (MLDv2) for IPv6", RFC 3810, DOI 10.17487/RFC3810, June 2004, <http://www.rfc-editor.org/info/rfc3810>.

[RFC4291] Hinden, R. and S. Deering, "IP Version 6 Addressing Architecture", RFC 4291, DOI 10.17487/RFC4291, February 2006, <http://www.rfc-editor.org/info/rfc4291>.

[RFC4541] Christensen, M., Kimball, K., and F. Solensky, "Considerations for Internet Group Management Protocol (IGMP) and Multicast Listener Discovery (MLD) Snooping Switches", RFC 4541, DOI 10.17487/RFC4541, May 2006, <http://www.rfc-editor.org/info/rfc4541>.

[RFC4861] Narten, T., Nordmark, E., Simpson, W., and H. Soliman, "Neighbor Discovery for IP version 6 (IPv6)", RFC 4861, DOI 10.17487/RFC4861, September 2007, <http://www.rfc-editor.org/info/rfc4861>.

[RFC6018] Baker, F., Harrop, W., and G. Armitage, "IPv4 and IPv6 Greynets", RFC 6018, DOI 10.17487/RFC6018, September 2010, <http://www.rfc-editor.org/info/rfc6018>.

[RFC6434] Jankiewicz, E., Loughney, J., and T. Narten, "IPv6 Node Requirements", RFC 6434, DOI 10.17487/RFC6434, December 2011, <http://www.rfc-editor.org/info/rfc6434>.

[RFC6583] Gashinsky, I., Jaeggli, J., and W. Kumari, "Operational Neighbor Discovery Problems", RFC 6583, DOI 10.17487/RFC6583, March 2012, <http://www.rfc-editor.org/info/rfc6583>.

[RFC6636] Asaeda, H., Liu, H., and Q. Wu, "Tuning the Behavior of the Internet Group Management Protocol (IGMP) and Multicast Listener Discovery (MLD) for Routers in Mobile and Wireless Networks", RFC 6636, DOI 10.17487/RFC6636, May 2012, <http://www.rfc-editor.org/info/rfc6636>.

[RFC6724] Thaler, D., Ed., Draves, R., Matsumoto, A., and T. Chown, "Default Address Selection for Internet Protocol Version 6 (IPv6)", RFC 6724, DOI 10.17487/RFC6724, September 2012, <http://www.rfc-editor.org/info/rfc6724>.

[RFC7217] Gont, F., "A Method for Generating Semantically Opaque Interface Identifiers with IPv6 Stateless Address Autoconfiguration (SLAAC)", RFC 7217, DOI 10.17487/RFC7217, April 2014, <http://www.rfc-editor.org/info/rfc7217>.

[RFC7421] Carpenter, B., Ed., Chown, T., Gont, F., Jiang, S., Petrescu, A., and A. Yourtchenko, "Analysis of the 64-bit Boundary in IPv6 Addressing", RFC 7421, DOI 10.17487/RFC7421, January 2015, <http://www.rfc-editor.org/info/rfc7421>.

Author’s Address

Mark Smith
PO BOX 521
HEIDELBERG, VIC 3084
AU

Email: markzzzsmith@gmail.com