Network Working Group
Internet-Draft
Intended status: Informational
Expires: September 2, 2019

K. Fujiwara
JPRS

March 01, 2019

Measures against cache poisoning attacks using IP fragmentation in DNS draft-fujiwara-dnsop-fragment-attack-01

Abstract

Researchers proposed practical DNS cache poisoning attacks using IP fragmentation. This document shows feasible and adequate measures at full-service resolvers and authoritative servers against these attacks. To protect resolvers from these attacks, avoid fragmentation (limit requestor's UDP payload size to 1220/1232), drop fragmented UDP DNS responses and use TCP at resolver side. To make a domain name robust against these attacks, limit EDNS0 Responder's maximum payload size to 1220, set DONTFRAG option to DNS response packets and use good random fragmentation ID at authoritative server side.

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 September 2, 2019.

Copyright Notice

Table of Contents

1. Introduction

"Fragmentation Considered Poisonous" [Herzberg2013] proposed effective off-path DNS cache poisoning attacks using IP fragmentation. The attacks mainly depend on the use of UDP to retrieve long DNS responses, resulting in packet fragmentation.

Recent full-service resolvers use good randomness for query source port numbers and ID field in DNS header to prevent cache poisoning attacks by off-path attackers. However, IP fragmentation is performed by OS kernel or routers that operators of DNS servers cannot control, and the query source port number and ID field in DNS header exist only in first fragment. The attack depends on poor randomness of "Identification" field generated by IP fragmentation and some bugs in IP reassembly code. Attackers can know path MTU size between authoritative servers and victim full-service resolvers, and responses from the authoritative servers. If attackers know generation algorithm of the "Identification" field, they can generate crafted second fragment packets that will be accepted by victim full-service resolvers.

[Hlavacek2013] also discussed the attacks and pointed that attackers can control path MTU size between some authoritative servers and victim full-service resolvers by sending crafted ICMP packets (Fragmentation needed and DF set, or ICMPv6 Packet Too Big). [Hlavacek2013] proposed a defense and two workarounds. The defence is DNSSEC and workarounds are ignoring ICMP type=3 code=4 (fragmentation needed and DF set), and limiting response size / EDNS0 buffer size fit to MTU size.

And more, "Domain Validation++ For MitM-Resilient PKI" [Brandt2018] proved that off-path attackers can intervene in path MTU discovery [RFC1191] to perform intentionally fragment responses from authoritative servers. They also proved that they poisoned Certificate Authorities (CAs)' full-service resolvers and successfully issued some fraudulent certificates.

As a result, we cannot trust all fragmented UDP packets and path MTU discovery.

By the way, TCP is considered strong against fragmentation attacks because TCP has sequence number and acknowledgement number in each sequence.

This document describes possible measures of cache poisoning attacks using IP fragmentation.

2. Methodology of the attack

DNS cache poisoning attacks using IP fragmentation are performed by combining the path MTU attack and cache poisoning attack. Path MTU attack targets are authoritative DNS servers. Cache poisoning attack targets are full-service resolvers.

Cache poisoning attacks using IP fragmentation are performed as follows steps. Path MTU attack is performed by step 3. Cache poisoning attack is performed by step 4 to 6.

1. Choose victim full-service resolver and target domain name.

2. Get the correct response from authoritative servers of the target domain name.

3. Send crafted ICMP/ICMPv6 packet to authoritative servers of the target domain name. The crafted ICMP packet indicates small path MTU size from the authoritative server to the victim full-resolver. If control of the path MTU succeed, proceed to the next step.

4. Generate second fragments from the correct response retrieved at step 2 with specified path MTU size, and calculate partial checksum value of the second fragment. Generate crafted second fragment that has the same partial checksum value. (If the partial checksum value of the correct second fragment and the partial checksum value of the crafted second fragment are the same, the UDP checksum value are the same.)

5. Send trigger query (target domain name / type) to the victim full-service resolver.

6. Send the crafted second fragment to victim full-service resolver with assumed fragment ID (or all possible IDs, at most 65536 on IPv4).

7. If victim full-service resolver accepts the crafted second fragment, the attack is successful.

The keys of the attack are:

- The attacker can control the fragmentation.

- The attacker can generate second fragment that generates the same UDP checksum value as the original response.

- The query source port and DNS ID field exist in the first fragment.

- the reassembly process holds received second fragment until arrival of the first fragment (timing is not strict),

- IPv4 fragmentation ID field has only 16 bits.

- Some IPv6 implementations use predictable fragment Identification values [RFC7739].

The probability of spoofing a resolver is described in Section 7.2 of [RFC5452]. The DNS cache poisoning attack using IP fragmentation changes to P=1 and I=1 (source port and ID are in the first fragment and need not predict), and adds number of fragment IDs as a denominator.

On IPv6, the attack does not change the probability because IPv6 fragment ID field has 32 bits. On IPv4, the attack changes the probability from 1/2^32 to 1/2^16 because IPv4 fragment ID field has only 16 bits.

3. Current status

[Brandt2018] showed that Linux version 3.13 and older versions are vulnerable to crafted ICMP fragmentation needed and DF set packet and off-path attackers can set some of authoritative servers' path MTU size to 296.

The author tested Linux version 2.6.32, 4.18.20 and FreeBSD 12.0. Linux 2.6.32 accepts crafted "ICMP Need Fragmentation and DF set" packet and path MTU decreased to 552. Linux 2.6.32, Linux 4.18.20 and FreeBSD 12.0 accept crafted "ICMPv6 Packet Too Big" packet and path MTU decreased to 1280.

Linux version 4.18.20 may ignore crafted ICMP packet.

FreeBSD and NetBSD accept "ICMP Need Fragmentation and DF set" packet related to established TCP and ignore "ICMP Need Fragmentation and DF set" packet related to UDP.

Then, off-path attackers can decrease path MTU size from some IPv4 authoritative servers to 552 (or 296), and can decrease path MTU size from IPv6 authoritative servers to 1280 (minimal IPv6 MTU size).

As described before, some old operating systems use predictable (incremental) fragmentation ID.

Furthermore, off-path attackers can know path MTU size related to authoritative servers and they can generate crafted fragmented DNS responses to victim full-service resolvers.

Then, measures against these attacks at full-service resolvers is important.

OS / source	crafted ICMPv4	minimal IPv4 MTU	crafted ICMPv6	minimal IPv6 MTU
[Brandt2018]	accept	552/296	unknown	unknown
Linux 2.6.32	accept	552	accept	1280
Linux 4.18.20	ignore?		accept	1280
FreeBSD 12	ignore		accept	1280

4. Possible measures

4.1. Use DNSSEC

DNSSEC is a measure against cache poisoning attacks. However, there are many unsigned zones and full-service resolver operator need to consider these zones.

"Use DNSSEC" requires both authoritative side and resolver side support.

4.2. Limit requestor's UDP payload size to 1220/1232 on IPv6

Limiting EDNS0 requestor's UDP payload size [RFC6891] to 1220/1232 on IPv6 is a measure of path MTU attacks on IPv6 because minimal MTU size of IPv6 is 1280 and most of implementations ignore ICMPv6 packet too big packets whose MTU size is smaller than 1280.

4.3. Limit requestor's UDP payload size to 512

Limiting EDNS0 requestor's UDP payload size [RFC6891] to 512 may be a measure of path MTU attacks.

However, since most of DNSSEC responses exceed 512 octets, limiting EDNS0 requestor's UDP payload size to 512 results truncated responses and resolvers need to retry queries by TCP. It always decreases name resolution performance.

And more, [Brandt2018] showed that off-path attackers can set some of authoritative servers' path MTU cache to 296. In this case, limiting EDNS0 payload size is not a measure.

Section 3 of [RFC4035] defines that A security-aware name server MUST support a message size of at least 1220 octets.

4.4. Set IP_DONTFRAG / IPv6 DONTFRAG at authoritative servers

It is a measure of authoritative server side.

4.5. Drop path MTU discovery or filter ICMP related to path MTU discovery

It is not a measure of resolver side. All authoritative servers need to be changed. Changing all authoritative servers is impossible. TCP requires path MTU discovery.

4.6. Drop all fragmented packets

To avoid the fragmentation attacks, "drop all fragmented packets" is one of the ideas. However, under path MTU discovery attacks, TCP packets may be fragmented and dropped. Then, "drop all fragmented UDP packets related to DNS" is the solution.

4.7. Drop fragmented UDP DNS responses at full-service resolvers

Drop fragmented UDP DNS responses at full-service resolvers may be a measure of cache poisoning attacks using IP fragmentation.

To avoid fragmentation in normal condition, use EDNS0 requestor's and responder's UDP payload size as 1220 to avoid fragmentation. 1220 is the minimal value defined by [RFC4035].

Under path MTU discovery attacks and cache poisoning attacks using IP fragmentation, UDP DNS response packets are fragmented and dropped and name resolution fails.

If resolver software retries by TCP, TCP is strong for fragmentation attacks and name resolution by TCP will success.

4.8. Use TCP only

It is believed that TCP is not vulnerable to fragmentation attacks. Unbound has "tcp-upstream" option that changes the upstream queries use TCP only for transport.

Some operators that support [RFC8078] said that they use TCP only for transport to avoid cache poisoning attacks.

The full-service resolvers of multiple CAs issuing domain validation (DV) certificates are required to withstand cache poisoning attacks, it is better to implement their full-service resolvers use TCP upstream queries only. Section 11.2 "DNS security" of [I-D.ietf-acme-acme] recommends that servers SHOULD perform DNS

queries over TCP, which provides better resistance to some forgery attacks than DNS over UDP.

4.9. Use good randomness for Fragmentation Identification field

See [RFC7739].

5. Proposal

To avoid cache poisoning attacks using IP fragmentation by full-service resolvers,

- Full-service resolvers set EDNS0 requestor's UDP payload size to 1220. (minimal size defined by [RFC4035])
- Full-service resolvers drop fragmented UDP responses related to DNS.
- Full-service resolvers may retry name resolution by TCP.
- (Full-service resolvers support DNSSEC validation.)

To make a domain name robust for cache poisoning attacks using IP fragmentation,

- Authoritative servers choose EDNS0 responder's maximum payload size limit to 1220 (to avoid IP fragmentation).
- Authoritative servers send DNS responses with IP_DONTFRAG / IPV6_DONTFRAG options.
- (Authoritative servers support DNSSEC and sign the domain name.)
- Authoritative servers and network devices use good randomness for fragmentation Identification field.

Exception: If authoritative servers and full-service resolvers are located beyond the link with the MTU value less than 1280, choose EDNS0 requestor's and responder's maximum payload size limit to the smallest link MTU value.

6. Example firewall configuration

Linux iptables support dropping first fragment with UDP source port 53 by using m32 module. Other first fragments that is not UDP, not source port 53 are not dropped. Second and following fragments should not be dropped because they may relate to other protocols.

Second fragments related to DNS will be dropped because their first fragments dropped.

```
iptables -t raw -A PREROUTING -m u32 --u32 \\
    "6&0xFFFF00FF=0x20000011&&18&0xffff=53" -j DROP

or  iptables -t raw -A PREROUTING -p udp -f -j DROP

ip6tables -A INPUT -p udp -m frag --fragfirst -m udp --sport 53 -j DROP
```

Other OSs may not handle first fragments. Then, drop all fragmented UDP packets.

On FreeBSD, 'ipfw' can drop all fragmented UDP packets (second fragments).

```
ipfw deny log udp from any to me in frag
```

7. IANA Considerations

This document has no IANA actions.

8. Security Considerations

Under path MTU discovery and fragmentation attacks, most full-service resolver software do not retry name resolution by TCP, name resolution related to attacks fails.

9. Acknowledgments

The author would like to specifically thank Mark Andrews and Daisuke HIGASHI.

10. Change History

10.1. 00

Initial version

10.2. 01

- Added Attack methodology
- Added measures at authoritative servers

11. References

11.1. Normative References

[RFC1191] Mogul, J. and S. Deering, "Path MTU discovery", RFC 1191, DOI 10.17487/RFC1191, November 1990, <https://www.rfc-editor.org/info/rfc1191>.

[RFC4035] Arends, R., Austein, R., Larson, M., Massey, D., and S. Rose, "Protocol Modifications for the DNS Security Extensions", RFC 4035, DOI 10.17487/RFC4035, March 2005, <https://www.rfc-editor.org/info/rfc4035>.

[RFC5452] Hubert, A. and R. van Mook, "Measures for Making DNS More Resilient against Forged Answers", RFC 5452, DOI 10.17487/RFC5452, January 2009, <https://www.rfc-editor.org/info/rfc5452>.

[RFC6891] Damas, J., Graff, M., and P. Vixie, "Extension Mechanisms for DNS (EDNS(0))", STD 75, RFC 6891, DOI 10.17487/RFC6891, April 2013, <https://www.rfc-editor.org/info/rfc6891>.

[RFC7739] Gont, F., "Security Implications of Predictable Fragment Identification Values", RFC 7739, DOI 10.17487/RFC7739, February 2016, <https://www.rfc-editor.org/info/rfc7739>.

11.2. Informative References

[Brandt2018] Brandt, M., Dai, T., Klein, A., Shulman, H., and M. Waidner, "Domain Validation++ For MitM-Resilient PKI", Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security , 2018.

[Herzberg2013] Herzberg, A. and H. Shulman, "Fragmentation Considered Poisonous", IEEE Conference on Communications and Network Security , 2013.

[Hlavacek2013] Hlavacek, T., "IP fragmentation attack on DNS", RIPE 67 Meeting , 2013, <https://ripe67.ripe.net/presentations/240-ipfragattack.pdf>.

[I-D.ietf-acme-acme]
Barnes, R., Hoffman-Andrews, J., McCarney, D., and J. Kasten, "Automatic Certificate Management Environment (ACME)", draft-ietf-acme-acme-18 (work in progress), December 2018.

[RFC8078] Gudmundsson, O. and P. Wouters, "Managing DS Records from the Parent via CDS/CDNSKEY", RFC 8078, DOI 10.17487/RFC8078, March 2017, <https://www.rfc-editor.org/info/rfc8078>.

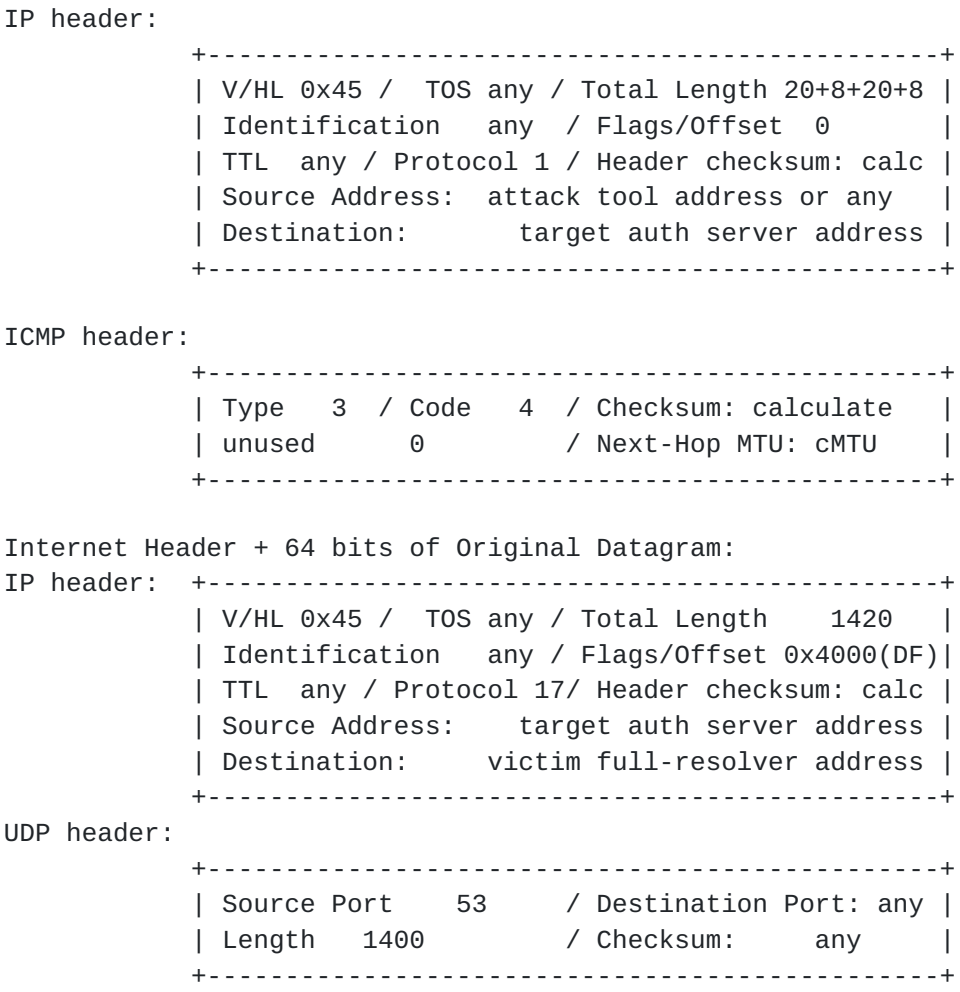
Appendix A. How to know path MTU size

- Linux: ip route get <IPv4/IPv6 address>
- FreeBSD: sysctl -o net.inet.tcp.hostcache.list

Appendix B. How to generate crafted ICMP packets

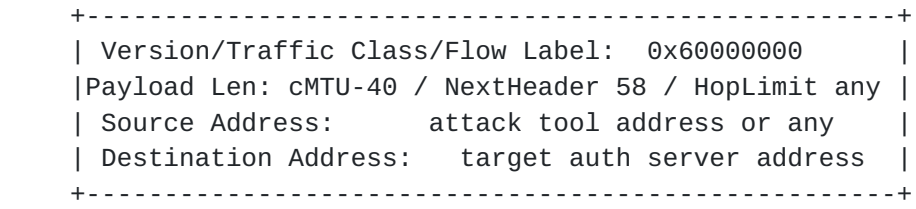
Let the crafted path MTU size be cMTU.

B.1. Example of crafted ICMP Need Fragmentation and DF set packet

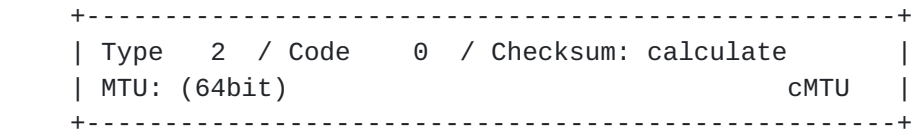


B.2. Example of crafted ICMPv6 Packet Too Big

IPv6 header:

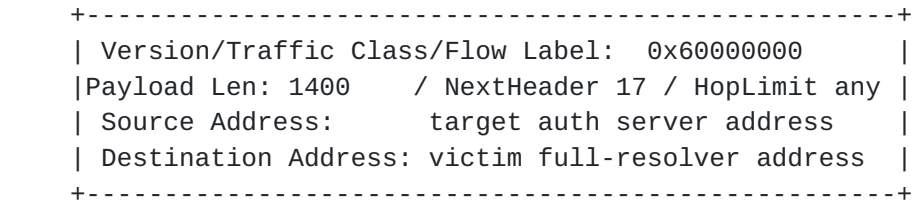


ICMPv6 header:



Fake invoking packet

IPv6 header:



UDP header:

Rest: Fill zero to end of packet

Author's Address

Kazunori Fujiwara
Japan Registry Services Co., Ltd.
Chiyoda First Bldg. East 13F, 3-8-1 Nishi-Kanda
Chiyoda-ku, Tokyo 101-0065
Japan

Phone: +81 3 5215 8451
Email: fujiwara@jprs.co.jp