draft-ietf-dnsop-avoid-fragmentation-05

K. Fujiwara, P. Vixie

dnsop WG at IETF 111

Updates from IETF 110

- Submitted draft-ietf-dnsop-avoid-fragmentation-05, June 23, 2021
  - Moved some text from Introduction to Appendix A. Weaknesses of IP fragmentation. (proposed by Brian Dickson)
  - Section 3.3: Default Maximum DNS/UDP payload size
    - Generated Table 1: Default maximum DNS/UDP payload size
    - 1400 is set as "Authors' recommendation"
    - Moved details to Appendix B. Details of maximum DNS/UDP payload size discussions.
    - Added new text: “Fragmentation avoidance is achieved with the IP(V6)_DONTFRAG option. The purpose of packet size limitation is to decrease packet loss due to the effects of the IP(V6)_DONTFRAG option.“
  - Added a new term: "IP_DONTFRAG"

Table 1 Default maximum DNS/UDP payload size

Source	IPv4	IPv6
RFC 4035 (MUST)	1220	1220
Software developers / DNSFlagDay2020 propose	1232	1232 (1280-40-8)
Authors' recommendation	1400	1400 (1500-40-8-some headers)
Maximum: Ethernet MTU 1500 [Huston2021]	1472 (1500-20-8)	1452 (1500-40-8)
Measured	MTU-20-8	MTU-40-8

- Operators MAY choose a good value from Table 1.

new term "IP_DONTFRAG"

- IP_DONTFRAG option is not defined by any RFCs.
- It is similar to IPV6_DONTFRAG option defined in [RFC3542].
- IP_DONTFRAG option is used on BSD systems to set the Don't Fragment bit [RFC0791] when sending IPv4 packets.
- On Linux systems this is done via IP_MTU_DISCOVER and IP_PMTUDISC_DO.

Current recommendations

3.1 Recommendations for UDP responders

- UDP responders SHOULD send DNS responses with IP_DONTFRAG / IPV6_DONTFRAG [RFC3542] options.
- (Fragmentation avoidance is achieved with the IP(V6)_DONTFRAG option. The purpose of packet size limitation is to decrease packet loss due to the effects of the IP(V6)_DONTFRAG option): choose good maximum DNS/UDP payload size

3.2 Recommendations for UDP requestors

- UDP requestors SHOULD send DNS requests with IP_DONTFRAG / IPV6_DONTFRAG [RFC3542] options.
- (Fragmentation avoidance is achieved with the IP(V6)_DONTFRAG option. The purpose of packet size limitation is to decrease packet loss due to the effects of the IP(V6)_DONTFRAG option): choose good maximum DNS/UDP payload size
- UDP requestors MAY drop fragmented DNS/UDP responses without IP reassembly to avoid cache poisoning attacks.
- DNS responses may be dropped by IP fragmentation. Upon a timeout, UDP requestors may retry using TCP or UDP, per local policy.

“DNS over TCP Considered Vulnerable”

- Haya Shulman et al. published a new paper
  - Tianxiang Dai, Haya Shulman, and Michael Waidner will present "DNS over TCP Considered Vulnerable” at ANRW 2021 (July 28, 2021)
    - See: https://irtf.org/anrw/2021/program.html
- They proposed that cache poisoning attacks using spoofed second fragment with valid IP_ID / checksum and forged payload for DNS over TCP
  1. Send ICMPv4 "fragmentation needed and DF set" to intermediate routers between the attack target resolver and authoritative servers
  2. Send a trigger query to the target resolver
  3. Estimate IP_ID, send spoofed second fragment to the target
     - The original TCP header (source port, seq/ack) exists in the first fragment.
     - Generating spoofed second fragment with the same checksum is easy (if the attacker know frag size and IP_ID)
- Attack targets
  - IPv4 (IPv6 may not be affected)
  - 496 of Alexa top-100K domains are vulnerable to fragmentation over TCP
- Please listen their presentation and evaluate the effect of fragmentation attack on DNS over TCP
- Possible solutions
  - RFC 6864 Updated Specification of the IPv4 ID Field
  - Setting IP_DF (on IPv4) for TCP (or, enable path MTU discovery on IPv4 TCP)

draft-ietf-dnsop-avoid-fragmentation-05

- Authors believe that the draft is ready for WGLC
- Please review carefully