Workgroup: Network Working Group
Internet-Draft: draft-hinden-v6ops-dns-00
Published: 3 May 2024
Intended Status: Best Current Practice
Expires: 4 November 2024
Authors: R. Hinden
Check Point Software
S. Krishnan
Cisco

DNS over IPv6 Best Practices

Abstract

This document describes an approach to how Domain Name Protocol (DNS) should be carried over IPv6. There have been some operational issues identified in carrying DNS packets over IPv6 and this draft proposes solutions to address them. A summary of what is proposed is to limit IPv6 DNS responses over UDP to be 1280 octets and use TCP or QUIC for anything larger.

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 4 November 2024.

Table of Contents

1. Introduction

This document describes an approach to how the Domain Name Protocol (DNS) should be carried over IPv6 [RFC8200]. There have been some operatonal issues identified in carrying DNS packets over IPv6 and this draft proposes solutions to address them.

The IPv6 protocol requires a minimum link MTU of 1280 octets. From Section 5 "Packet Size Issues" of RFC8200:

> IPv6 requires that every link in the Internet have an MTU of 1280 octets or greater. This is known as the IPv6 minimum link MTU. On any link that cannot convey a 1280-octet packet in one piece, link- specific fragmentation and reassembly must be provided at a layer below IPv6.

Using Path MTU Discovery and IPv6 fragmentation larger packets may be sent. However, there is operational experience that seems to indicate that sending large DNS over UDP over IPv6 packets results in high loss rates.

Most modern transport protocols like TCP [TCP] and [QUIC] include packet segmentation techniques that allow them to send larger data streams over IPv6.

A recent dnsop working group document titled "IP Fragmentation Avoidance in DNS over UDP" [I-D.ietf-dnsop-avoid-fragmentation] also describes the issue, and recommends as best current practice to disable IPv6 fragmentation for sending DNS packets over IPv6. Specifically:

> 3.1. Recommendations for UDP responders
>
> R1. UDP responders SHOULD NOT use IPv6 fragmentation [RFC8200].

This document is aligned with the recommendation in [I-D.ietf-dnsop-avoid-fragmentation], but focuses on DNS over IPv6, and also recommends and provides additional details on running DNS over TCP or QUIC.

2. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

3. Domain Name System Protocols

The Domain Name System was originally defined in [RFC1034] and [RFC1035]. It was designed to be able to run over different transport protocols, including UDP, and TCP. It has also recently been extended to be able to run over QUIC [RFC9250]. These transport protocols can be run over IPv4 or IPv6.

When DNS was originally designed, the size of the DNS packets carried over UDP was limited to 512 bytes as defined in Section 2.3.4. of [RFC1035]. Longer messages were truncated and the Truncation (TC) bit was set indicating the response is incomplete so that the client can retry with TCP.

With this original behavior UDP over IPv6 would not exceed the IPv6 link MTU and hence would not present operational issues due to fragmentation. When DNSSEC was introduced, several responses were too small to fit into this 512 byte limit for UDP and hence resulted in TCP connections which impacted the scalability of the DNS servers. [RFC6891] allowed the requestors to specify larger UDP message sizes that they are able to handle but the sizes specified by the requestor might exceed the Path MTU.

4. DNS over IPv6

DNS over IPv6 is defined to run over UDP, a very simple transport protocol, or more capable transport protocols like TCP or QUIC. UDP only provides for source and destination ports, a length field, and simple checksum. It is connectionless. Transport protocols like TCP and UDP provide additional features like packet segmentation, reliability, error correction, and connection state.

DNS over UDP over IPv6 works fine with small packet sizes, but becomes less reliable with larger packet sizes, especially in cases that require IPv6 datagram fragmentation.

DNS over TCP/QUIC over IPv6 will work fine with all packet sizes. The downside of running DNS over a stateful protocol such as TCP or QUIC is that this will require more resources on the DNS server and potentially affect scalability. This might be a reasonable tradeoff in case of servers that need to send larger DNS response packets.

4.1. DNS over UDP

It is recommended to limit packet sizes of DNS over UDP over IPv6 to 1280 octets. This avoids any need for IPv6 fragmentation or Path MTU Discovery. This should be very reliable.

Many (if not most) DNS queries and responses will fit with this packet size limit and hence can be sent over UDP. Larger DNS packets SHOULD not be sent over UDP, instead they SHOULD be sent over TCP or QUIC as described in the next section.

This recommendation is consistent with "UDP Usage Guidelines" [RFC8085] that makes the following recommendation when effective PMTU/PLPMTUD in not supported (as it is potentially the case in Internet-wide DNS operations):

> Applications that do not follow the recommendation to do PMTU/PLPMTUD discovery SHOULD still avoid sending UDP datagrams that would result in IP packets that exceed the path MTU. Because the actual path MTU is unknown, such applications SHOULD fall back to sending messages that are shorter than the default effective MTU for sending (EMTU_S in [RFC1122]). For IPv4, EMTU_S is the smaller of 576 bytes and the first-hop MTU [RFC1122]. For IPv6, EMTU_S is 1280 bytes [RFC2460].

4.2. DNS over TCP and QUIC

When larger DNS packets need to be carried, it is recommended to run DNS over TCP or QUIC. These protocols handle segmentation and will reliably adjust their segment size for different link and path MTU values. In this regard they work much more reliably than using UDP with IPv6 fragmentation.

Section 4.2.2. of [RFC1035] describes the of TCP for carrying DNS messages. [RFC9250] describes how to use DNS over QUIC in order to provide transport confidentiality. Operation requirements for DNS over TCP are described in [RFC9210]

5. IANA Considerations

There are no actions required for IANA defined in this document.

6. Security Considerations

Switching from UDP to TCP/QUIC for large responses means that the DNS server needs to maintain additional state for each query that was received over TCP/QUIC. This will consume additional resources on the servers and impact the scalability of the DNS system. It may also leave the servers vulnerable to DoS attacks.

7. Acknowledgments

Geoff Huston discussed the operational issues in his article "IPv6, DNS, and truncation in UDP" IPv6" [HUSTON] and Jared Mauch brought up the issues to the 6MAN working group.

Helpful comments were received from Warren Kumari, and other members of the V6OPS working group.

8. Change log [RFC Editor: Please remove]

draft-hinden-ipv6-v6ops-00, 2024-May-02:

*Initial draft.

9. Normative References

[RFC1034] Mockapetris, P., "Domain names - concepts and facilities", STD 13, RFC 1034, DOI 10.17487/RFC1034, November 1987, <https://www.rfc-editor.org/info/rfc1034>.

[RFC1035] Mockapetris, P., "Domain names - implementation and specification", STD 13, RFC 1035, DOI 10.17487/RFC1035, November 1987, <https://www.rfc-editor.org/info/rfc1035>.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC8085] Eggert, L., Fairhurst, G., and G. Shepherd, "UDP Usage Guidelines", BCP 145, RFC 8085, DOI 10.17487/RFC8085, March 2017, <https://www.rfc-editor.org/info/rfc8085>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

[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/info/rfc8200>.

[RFC9210] Kristoff, J. and D. Wessels, "DNS Transport over TCP - Operational Requirements", BCP 235, RFC 9210, DOI 10.17487/RFC9210, March 2022, <https://www.rfc-editor.org/info/rfc9210>.

[RFC9250] Huitema, C., Dickinson, S., and A. Mankin, "DNS over Dedicated QUIC Connections", RFC 9250, DOI 10.17487/RFC9250, May 2022, <https://www.rfc-editor.org/info/rfc9250>.

10. Informative References

[HUSTON] Huston, G., "IPv6, DNS, and truncation in UDP", , , March 2024, <https://blog.apnic.net/2024/03/14/ipv6-dns-and-truncation-in-udp/>.

[I-D.ietf-dnsop-avoid-fragmentation] Fujiwara, K. and P. A. Vixie, "IP Fragmentation Avoidance in DNS over UDP", Work in Progress, Internet-Draft, draft-ietf-dnsop-avoid-fragmentation-17, 29 February 2024, <https://datatracker.ietf.org/doc/html/draft-ietf-dnsop-avoid-fragmentation-17>.

[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>.

Authors' Addresses

Robert M. Hinden
Check Point Software
100 Oracle Parkway, Suite 800
Redwood City, CA 94065
United States of America

Email: bob.hinden@gmail.com

Suresh Krishnan
Cisco
United States of America

Email: suresh.krishnan@gmail.com