

Domain Name System Operations  
Internet-Draft  
Intended status: Best Current Practice  
Expires: December 22, 2017

J. Kristoff  
DePaul University  
D. Wessels  
Verisign  
June 20, 2017

**DNS Transport over TCP - Operational Requirements**  
**draft-ietf-dnsop-dns-tcp-requirements-00**

Abstract

This document encourages the practice of permitting DNS messages to be carried over TCP on the Internet. It also describes some of the consequences of this behavior and the potential operational issues that can arise when this best common practice is not upheld.

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 December 22, 2017.

Copyright Notice

Copyright (c) 2017 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

## Table of Contents

<a href="#">1.</a>	Introduction . . . . .	<a href="#">2</a>
<a href="#">1.1.</a>	Requirements Language . . . . .	<a href="#">3</a>
<a href="#">2.</a>	Background . . . . .	<a href="#">3</a>
<a href="#">2.1.</a>	Uneven Transport Usage and Preference . . . . .	<a href="#">3</a>
<a href="#">2.2.</a>	Waiting for Large Messages and Reliability . . . . .	<a href="#">3</a>
<a href="#">2.3.</a>	EDNS0 . . . . .	<a href="#">4</a>
<a href="#">2.4.</a>	"Only Zone Transfers Use TCP" . . . . .	<a href="#">5</a>
<a href="#">3.</a>	DNS over TCP Requirements . . . . .	<a href="#">5</a>
<a href="#">4.</a>	Network and System Considerations . . . . .	<a href="#">6</a>
<a href="#">5.</a>	DNS over TCP Filtering Risks . . . . .	<a href="#">6</a>
<a href="#">5.1.</a>	DNS Wedgie . . . . .	<a href="#">6</a>
<a href="#">5.2.</a>	DNS Root Zone KSK Rollover . . . . .	<a href="#">6</a>
<a href="#">6.</a>	Acknowledgements . . . . .	<a href="#">7</a>
<a href="#">7.</a>	IANA Considerations . . . . .	<a href="#">7</a>
<a href="#">8.</a>	Security Considerations . . . . .	<a href="#">7</a>
<a href="#">9.</a>	References . . . . .	<a href="#">8</a>
<a href="#">9.1.</a>	Normative References . . . . .	<a href="#">8</a>
<a href="#">9.2.</a>	Informative References . . . . .	<a href="#">8</a>
<a href="#">Appendix A.</a>	Standards Related to DNS Transport over TCP . . . . .	<a href="#">10</a>
<a href="#">A.1.</a>	TODO - additional, relevant RFCs . . . . .	<a href="#">10</a>
<a href="#">A.2.</a>	IETF <a href="#">RFC 7477</a> - Child-to-Parent Synchronization in DNS . . . . .	<a href="#">11</a>
<a href="#">A.3.</a>	IETF <a href="#">RFC 7766</a> - DNS Transport over TCP - Implementation Requirements . . . . .	<a href="#">11</a>
<a href="#">A.4.</a>	IETF <a href="#">RFC 7828</a> - The edns-tcp-keepalive EDNS0 Option . . . . .	<a href="#">11</a>
<a href="#">A.5.</a>	IETF <a href="#">RFC 7873</a> - Domain Name System (DNS) Cookies . . . . .	<a href="#">11</a>
<a href="#">A.6.</a>	IETF <a href="#">RFC 7901</a> - CHAIN Query Requests in DNS . . . . .	<a href="#">11</a>
<a href="#">A.7.</a>	IETF <a href="#">RFC 8027</a> - DNSSEC Roadblock Avoidance . . . . .	<a href="#">12</a>
	Authors' Addresses . . . . .	<a href="#">12</a>

## [1.](#) Introduction

DNS messages may be delivered using UDP or TCP communications. While most DNS transactions are carried over UDP, some operators have been led to believe that any DNS over TCP traffic is unwanted or unnecessary for general DNS operation. As usage and features have evolved, TCP transport has become increasingly important for correct and safe operation of the Internet DNS. Reflecting modern usage, the DNS standards were recently updated to declare support for TCP is now a required part of the DNS implementation specifications in [\[RFC7766\]](#). This document is the formal requirements equivalent for the operational community, encouraging operators to ensure DNS over TCP communications support is on par with DNS over UDP communications.



### **1.1. Requirements Language**

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC 2119](#) [[RFC2119](#)].

## **2. Background**

### **2.1. Uneven Transport Usage and Preference**

In the original suite of DNS specifications, [[RFC1034](#)] and [[RFC1035](#)] clearly specified that DNS messages could be carried in either UDP or TCP, but they also made clear a preference for UDP as the transport for queries in the general case. As stated in [[RFC1035](#)]:

"While virtual circuits can be used for any DNS activity, datagrams are preferred for queries due to their lower overhead and better performance."

Another early, important, and influential document, [[RFC1123](#)], detailed the preference for UDP more explicitly:

"DNS resolvers and recursive servers MUST support UDP, and SHOULD support TCP, for sending (non-zone-transfer) queries."

and further stipulated:

A name server MAY limit the resources it devotes to TCP queries, but it SHOULD NOT refuse to service a TCP query just because it would have succeeded with UDP.

Culminating in [[RFC1536](#)], DNS over TCP came to be associated primarily with the zone transfer mechanism, while most DNS queries and responses were seen as the dominion of UDP.

### **2.2. Waiting for Large Messages and Reliability**

As stipulated in the original specifications, DNS messages over UDP were restricted to a 512-byte message size. However, even while [[RFC1123](#)] made a clear preference for UDP, it foresaw DNS over TCP becoming more popular in the future:

"[...] it is also clear that some new DNS record types defined in the future will contain information exceeding the 512 byte limit that applies to UDP, and hence will require TCP."

At least two new, widely anticipated developments were set to elevate the need for DNS over TCP transactions. The first was dynamic



updates defined in [[RFC2136](#)] and the second was the set of extensions collectively known as DNSSEC originally specified in [[RFC2541](#)]. The former suggested "requestors who require an accurate response code must use TCP", while the later warned "[...] larger keys increase the size of KEY and SIG RRs. This increases the chance of DNS UDP packet overflow and the possible necessity for using higher overhead TCP in responses."

Yet defying some expectations, DNS over TCP remained little used in real traffic across the Internet. Dynamic updates saw little deployment between autonomous networks. Around the time DNSSEC was first defined, another new feature affecting DNS over UDP helped solidify its dominance for message transactions.

### **[2.3.](#) EDNS0**

In 1999 the IETF published the Extension Mechanisms for DNS (EDNS0) in [[RFC2671](#)]. This document standardized a way for communicating DNS nodes to perform rudimentary capabilities negotiation. One such capability written into the base specification and present in every EDNS0 compatible message is the value of the maximum UDP payload size the sender can support. This unsigned 16-bit field specifies in bytes the maximum DNS MTU. In practice, typical values are a subset of the 512 to 4096 byte range. EDNS0 was rapidly and widely deployed over the next several years and numerous surveys have shown many systems currently support larger UDP MTUs [[CASTRO2010](#)], [[NETALYZR](#)] with EDNS0.

The natural effect of EDNS0 deployment meant large DNS messages would be less reliant on TCP than they might otherwise have been. While a nonnegligible population of DNS systems lack EDNS0 or may still fall back to TCP for some transactions, DNS over TCP transactions remain a very small fraction of overall DNS traffic [[VERISIGN](#)]. Nevertheless, some average increase in DNS message size, the continued development of new DNS features and a denial of service mitigation technique (see [Section 8](#)) have suggested that DNS over TCP transactions are as important to the correct and safe operation of the Internet DNS as ever, if not more so. Furthermore, there has been serious research that has suggested connection-oriented DNS transactions may provide security and privacy advantages over UDP transport [[TDNS](#)]. In fact, [[RFC7858](#)], a Standards Track document is just this sort of specification. Therefore, it might be desirable for network operators to avoid artificially inhibiting the potential utility and advances in the DNS such as these.



#### **2.4. "Only Zone Transfers Use TCP"**

There are many in the DNS community who configure DNS over TCP services and expect DNS over TCP transactions to occur without interference. However there has also been a long held belief by some operators, particularly for security-related reasons, that DNS over TCP services should be purposely limited or not provided at all [CHES94], [DJBDNS]. A popular meme has also held the imagination of some that DNS over TCP is only ever used for zone transfers and is generally unnecessary otherwise, with filtering all DNS over TCP traffic even described as a best practice.

The position on restricting DNS over TCP had some justification given that historic implementations of DNS nameservers provided very little in the way of TCP connection management (for example see [Section 6.1.2 of \[RFC7766\]](#) for more details). However modern standards and implementations are moving to align with the more sophisticated TCP management techniques employed by, for example, HTTP(S) servers and load balancers.

### **3. DNS over TCP Requirements**

[Section 6.1.3.2 in \[RFC1123\]](#) is updated: All general-purpose DNS servers MUST be able to service both UDP and TCP queries.

- o Authoritative servers MUST service TCP queries so that they do not limit the size of responses to what fits in a single UDP packet.
- o Recursive servers (or forwarders) MUST service TCP queries so that they do not prevent large responses from a TCP-capable server from reaching its TCP-capable clients.

Regarding the choice of limiting the resources a server devotes to queries, [Section 6.1.3.2 in \[RFC1123\]](#) also says:

A name server MAY limit the resources it devotes to TCP queries, but it SHOULD NOT refuse to service a TCP query just because it would have succeeded with UDP.

This requirement is hereby updated: A name server MAY limit the the resources it devotes to queries, but it MUST NOT refuse to service a query just because it would have succeeded with another transport protocol.

DNS over TCP filtering is considered harmful in the general case. DNS resolver and server operators MUST provide DNS service over both UDP and TCP transports. Likewise, network operators MUST allow DNS service over both UDP and TCP transports. It must be acknowledged



that DNS over TCP service can pose operational challenges that are not present when running DNS over UDP alone. However, it is the aim of this document to argue that the potential damage incurred by prohibiting DNS over TCP service is more detrimental to the continued utility and success of the DNS than when its usage is allowed.

#### **4. Network and System Considerations**

TODO: refer to IETF [RFC 7766](#) connection handling discussion, various TCP hardening documents, network operator protocol and traffic best practices, etc.

#### **5. DNS over TCP Filtering Risks**

Networks that filter DNS over TCP risk losing access to significant or important pieces of the DNS name space. For a variety of reasons a DNS answer may require a DNS over TCP query. This may include large message sizes, lack of EDNS0 support, DDoS mitigation techniques, or perhaps some future capability that is as yet unforeseen will also demand TCP transport.

Even if any or all particular answers have consistently been returned successfully with UDP in the past, this continued behavior cannot be guaranteed when DNS messages are exchanged between autonomous systems. Therefore, filtering of DNS over TCP is considered harmful and contrary to the safe and successful operation of the Internet. This section enumerates some of the known risks we know about at the time of this writing when networks filter DNS over TCP.

##### **5.1. DNS Wedgie**

Networks that filter DNS over TCP may inadvertently cause problems for third party resolvers as experienced by [\[TOYAMA\]](#). If for instance a resolver receives a truncated answer from a server, but if when the resolver resends the query using TCP and the TCP response never arrives, not only will full answer be unavailable, but the resolver will incur the full extent of TCP retransmissions and time outs. This situation might place extreme strain on resolver resources. If the number and frequency of these truncated answers are sufficiently high, we refer to the steady-state of lost resources as a result a "DNS" wedgie". A DNS wedgie is often not easily or completely mitigated by the affected DNS resolver operator.

##### **5.2. DNS Root Zone KSK Rollover**

Recent plans for a new root zone DNSSEC KSK have highlighted a potential problem in retrieving the keys.[\[LEWIS\]](#) Some packets in the KSK rollover process will be larger than 1280 bytes, the IPv6 minimum



MTU for links carrying IPv6 traffic.[\[RFC2460\]](#) While studies have shown that problems due to fragment filtering or an inability to generate and receive these larger messages are negligible, any DNS server that is unable to receive large DNS over UDP messages or perform DNS over TCP may experience severe disruption of DNS service if performing DNSSEC validation.

## **6. Acknowledgements**

This document was initially motivated by feedback from students who pointed out that they were hearing contradictory information about filtering DNS over TCP messages. Thanks in particular to a teaching colleague, JPL, who perhaps unknowingly encouraged the initial research into the differences of what the community has historically said and did. Thanks to all the NANOG 63 attendees who provided feedback to an early talk on this subject.

The following individuals provided an array of feedback to help improve this document: Sara Dickinson, Bob Harold, Tatuya Jinmei, and Paul Hoffman. The authors are indebted to their contributions. Any remaining errors or imperfections are the sole responsibility of the document authors.

## **7. IANA Considerations**

This memo includes no request to IANA.

## **8. Security Considerations**

Ironically, returning truncated DNS over UDP answers in order to induce a client query to switch to DNS over TCP has become a common response to source address spoofed, DNS denial-of-service attacks [\[RRL\]](#). Historically, operators have been wary of TCP-based attacks, but in recent years, UDP-based flooding attacks have proven to be the most common protocol attack on the DNS. Nevertheless, a high rate of short-lived DNS transactions over TCP may pose challenges. While many operators have provided DNS over TCP service for many years without duress, past experience is no guarantee of future success.

DNS over TCP is not unlike many other Internet TCP services. TCP threats and many mitigation strategies have been well documented in a series of documents such as [\[RFC4953\]](#), [\[RFC4987\]](#), [\[RFC5927\]](#), and [\[RFC5961\]](#).



## **9. References**

### **9.1. Normative References**

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

### **9.2. Informative References**

- [CASTRO2010] Castro, S., Zhang, M., John, W., Wessels, D., and k. claffy, "Understanding and preparing for DNS evolution", 2010.
- [CHES94] Cheswick, W. and S. Bellovin, "Firewalls and Internet Security: Repelling the Wily Hacker", 1994.
- [DJBDNS] D.J. Bernstein, "When are TCP queries sent?", 2002, <<https://cr.yp.to/djbdns/tcp.html#why>>.
- [LEWIS] Lewis, E., "2017 DNSSEC KSK Rollover", RIPE 74 Budapest, Hungary, May 2017.
- [NETALYZR] Kreibich, C., Weaver, N., Nechaev, B., and V. Paxson, "Netalyzer: Illuminating The Edge Network", 2010.
- [RFC1034] Mockapetris, P., "Domain names - concepts and facilities", STD 13, [RFC 1034](#), DOI 10.17487/RFC1034, November 1987, <<http://www.rfc-editor.org/info/rfc1034>>.
- [RFC1035] Mockapetris, P., "Domain names - implementation and specification", STD 13, [RFC 1035](#), DOI 10.17487/RFC1035, November 1987, <<http://www.rfc-editor.org/info/rfc1035>>.
- [RFC1123] Braden, R., Ed., "Requirements for Internet Hosts - Application and Support", STD 3, [RFC 1123](#), DOI 10.17487/RFC1123, October 1989, <<http://www.rfc-editor.org/info/rfc1123>>.
- [RFC1536] Kumar, A., Postel, J., Neuman, C., Danzig, P., and S. Miller, "Common DNS Implementation Errors and Suggested Fixes", [RFC 1536](#), DOI 10.17487/RFC1536, October 1993, <<http://www.rfc-editor.org/info/rfc1536>>.



- [RFC2136] Vixie, P., Ed., Thomson, S., Rekhter, Y., and J. Bound, "Dynamic Updates in the Domain Name System (DNS UPDATE)", [RFC 2136](#), DOI 10.17487/RFC2136, April 1997, <<http://www.rfc-editor.org/info/rfc2136>>.
- [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>>.
- [RFC2541] Eastlake 3rd, D., "DNS Security Operational Considerations", [RFC 2541](#), DOI 10.17487/RFC2541, March 1999, <<http://www.rfc-editor.org/info/rfc2541>>.
- [RFC2671] Vixie, P., "Extension Mechanisms for DNS (EDNS0)", [RFC 2671](#), DOI 10.17487/RFC2671, August 1999, <<http://www.rfc-editor.org/info/rfc2671>>.
- [RFC4953] Touch, J., "Defending TCP Against Spoofing Attacks", [RFC 4953](#), DOI 10.17487/RFC4953, July 2007, <<http://www.rfc-editor.org/info/rfc4953>>.
- [RFC4987] Eddy, W., "TCP SYN Flooding Attacks and Common Mitigations", [RFC 4987](#), DOI 10.17487/RFC4987, August 2007, <<http://www.rfc-editor.org/info/rfc4987>>.
- [RFC5927] Gont, F., "ICMP Attacks against TCP", [RFC 5927](#), DOI 10.17487/RFC5927, July 2010, <<http://www.rfc-editor.org/info/rfc5927>>.
- [RFC5961] Ramaiah, A., Stewart, R., and M. Dalal, "Improving TCP's Robustness to Blind In-Window Attacks", [RFC 5961](#), DOI 10.17487/RFC5961, August 2010, <<http://www.rfc-editor.org/info/rfc5961>>.
- [RFC7477] Hardaker, W., "Child-to-Parent Synchronization in DNS", [RFC 7477](#), DOI 10.17487/RFC7477, March 2015, <<http://www.rfc-editor.org/info/rfc7477>>.
- [RFC7766] Dickinson, J., Dickinson, S., Bellis, R., Mankin, A., and D. Wessels, "DNS Transport over TCP - Implementation Requirements", [RFC 7766](#), DOI 10.17487/RFC7766, March 2016, <<http://www.rfc-editor.org/info/rfc7766>>.
- [RFC7828] Wouters, P., Abley, J., Dickinson, S., and R. Bellis, "The edns-tcp-keepalive EDNS0 Option", [RFC 7828](#), DOI 10.17487/RFC7828, April 2016, <<http://www.rfc-editor.org/info/rfc7828>>.



- [RFC7858] Hu, Z., Zhu, L., Heidemann, J., Mankin, A., Wessels, D., and P. Hoffman, "Specification for DNS over Transport Layer Security (TLS)", [RFC 7858](#), DOI 10.17487/RFC7858, May 2016, <<http://www.rfc-editor.org/info/rfc7858>>.
- [RFC7873] Eastlake 3rd, D. and M. Andrews, "Domain Name System (DNS) Cookies", [RFC 7873](#), DOI 10.17487/RFC7873, May 2016, <<http://www.rfc-editor.org/info/rfc7873>>.
- [RFC7901] Wouters, P., "CHAIN Query Requests in DNS", [RFC 7901](#), DOI 10.17487/RFC7901, June 2016, <<http://www.rfc-editor.org/info/rfc7901>>.
- [RFC8027] Hardaker, W., Gudmundsson, O., and S. Krishnaswamy, "DNSSEC Roadblock Avoidance", [BCP 207](#), [RFC 8027](#), DOI 10.17487/RFC8027, November 2016, <<http://www.rfc-editor.org/info/rfc8027>>.
- [RRL] Vixie, P. and V. Schryver, "DNS Response Rate Limiting (DNS RRL)", ISC-TN 2012-1 Draft1, April 2012.
- [TDNS] Zhu, L., Heidemann, J., Wessels, D., Mankin, A., and N. Somaiya, "Connection-oriented DNS to Improve Privacy and Security", 2015.
- [TOYAMA] Toyama, K., Ishibashi, K., Ishino, M., Yoshimura, C., and K. Fujiwara, "DNS Anomalies and Their Impacts on DNS Cache Servers", NANOG 32 Reston, VA USA, 2004.
- [VERISIGN] Thomas, M. and D. Wessels, "An Analysis of TCP Traffic in Root Server DITL Data", DNS-OARC 2014 Fall Workshop Los Angeles, 2014.

## **Appendix A. Standards Related to DNS Transport over TCP**

This section enumerates all known IETF RFC documents that are currently of status standard, informational, best common practice or experimental and either implicitly or explicitly make assumptions or statements about the use of TCP as a transport for the DNS germane to this document.

### **A.1. TODO - additional, relevant RFCs**



#### **[A.2.](#) IETF [RFC 7477](#) - Child-to-Parent Synchronization in DNS**

This standards track document [[RFC7477](#)] specifies a RRType and protocol to signal and synchronize NS, A, and AAAA resource record changes from a child to parent zone. Since this protocol may require multiple requests and responses, it recommends utilizing DNS over TCP to ensure the conversation takes place between a consistent pair of end nodes.

#### **[A.3.](#) IETF [RFC 7766](#) - DNS Transport over TCP - Implementation Requirements**

The standards track document [[RFC7766](#)] might be considered the direct ancestor of this operational requirements document. The implementation requirements document codifies mandatory support for DNS over TCP in compliant DNS software.

#### **[A.4.](#) IETF [RFC 7828](#) - The edns-tcp-keepalive EDNS0 Option**

This standards track document [[RFC7828](#)] defines an EDNS0 option to negotiate an idle timeout value for long-lived DNS over TCP connections. Consequently, this document is only applicable and relevant to DNS over TCP sessions and between implementations that support this option.

#### **[A.5.](#) IETF [RFC 7873](#) - Domain Name System (DNS) Cookies**

This standards track document [[RFC7873](#)] describes an EDNS0 option to provide additional protection against query and answer forgery. This specification mentions DNS over TCP as a reasonable fallback mechanism when DNS Cookies are not available. The specification does make mention of DNS over TCP processing in two specific situations. In one, when a server receives only a client cookie in a request, the server should consider whether the request arrived over TCP and if so, it should consider accepting TCP as sufficient to authenticate the request and respond accordingly. In another, when a client receives a BADCOOKIE reply using a fresh server cookie, the client should retry using TCP as the transport.

#### **[A.6.](#) IETF [RFC 7901](#) - CHAIN Query Requests in DNS**

This experimental specification [[RFC7901](#)] describes an EDNS0 option that can be used by a security-aware validating resolver to request and obtain a complete DNSSEC validation path for any single query. This document requires the use of DNS over TCP or a source IP address verified transport mechanism such as EDNS-COOKIE.[[RFC7873](#)]



**A.7. IETF [RFC 8027](#) - DNSSEC Roadblock Avoidance**

This document [[RFC8027](#)] details observed problems with DNSSEC deployment and mitigation techniques. Network traffic blocking and restrictions, including DNS over TCP messages, are highlighted as one reason for DNSSEC deployment issues. While this document suggests these sorts of problems are due to "non-compliant infrastructure" and is of type BCP, the scope of the document is limited to detection and mitigation techniques to avoid so-called DNSSEC roadblocks.

**Authors' Addresses**

John Kristoff  
DePaul University  
Chicago, IL 60604  
US

Phone: +1 312 493 0305  
Email: [jtk@depaul.edu](mailto:jtk@depaul.edu)  
URI: <https://aharp.iorc.depaul.edu>

Duane Wessels  
Verisign  
12061 Bluemont Way  
Reston, VA 20190  
US

Phone: +1 703 948 3200  
Email: [dwessels@verisign.com](mailto:dwessels@verisign.com)  
URI: <http://verisigninc.com>

