Recommendations for dealing with fragmentation in DNS(SEC)

Abstract

DNS response messages can sometimes be large enough to exceed the Maximum Transmission Unit (MTU) size for the underlying physical network. These tend to be for responses containing large RR sets, or DNSSEC-related data. Large DNS responses can experience link-level fragmentation and may subsequently be blocked by firewalls, which reject fragmented IP packets. Resolving name servers behind such firewalls will therefore be unable to receive the complete response message. This can have undesirable results, such as certain DNS zones appearing to be unreachable. Fragmentation of DNS responses can be avoided by limiting the size of response messages. This can be done in the resolving name server, or at the authoritative name server. This document suggests response size limitations at authoritative name servers that could be applied until recommendations for resolving name servers are standardised and implemented.

1. Conventions used in this document

Domain names used in this document are for explanatory purposes only and should not be expected to lead to useful information in real life [RFC 2606].

The keywords MUST, MUST NOT, REQUIRED, SHALL, SHALL NOT, SHOULD, SHOULD NOT, RECOMMENDED, MAY, and OPTIONAL, when they appear in this document, are to be interpreted as described in [RFC 2119].

2. Background

2.1 Introduction

The maximum response size for traditional DNS is 512 bytes. The use of EDNS0 [RFC 2671] in DNS queries and responses allows for larger messages. Large response may be fragmented because they exceed the Maximum Transmission Unit (MTU) size of the underlying transport medium. By default the maximum response size advertised in queries with EDNS0 is 4096 bytes [RFC 2671BIS (DRAFT), Sec. 6]. Considering that most path MTUs are the MTU of Ethernet of 1500 bytes, this can result in fragmented responses.

For an MTU of 1500 bytes, the maximum size of an unfragmented response is 1472 bytes for IPv4. The remaining 28 bytes are used for a 20 byte IPv4 packet header (assuming no options were used) and an 8 byte UDP header. The minimum MTU for IPv6 is 1280 bytes [RFC2460]. Therefore the maximum payload for an unfragmented response over IPv6 is 1232 bytes after allowing 40 bytes for the IPv6 packet header and 8 bytes for the UDP header.

Some firewalls are set to block fragmented UDP messages due to past vulnerabilities (see also [NIST1]), as a result, fragmented responses sent to resolving name servers behind such firewalls will be blocked[1]. Also, DNS proxies and forwarders (for example, those included in CPE devices) may experience problems handling fragmented UDP messages.

2.2 Solution

Limiting the maximum response size can avoid fragmentation and hence improve delivery of responses. This can be done by a resolving name server advertising a maximum response size in queries with a value that avoids most fragmentation. Alternatively, an authoritative name server can be set to limit its response size.

2.3 Recommendation scope

Firewalls between resolving name servers and authoritative name servers should adhere to recent best practice guidelines that suggest not to block fragmented UDP messages [NIST1, Sec. 2.1.1]. In practice, however, this may not be possible (due to auditors requiring implementation of out-dated practices). In addition to that, research [DNSRESEARCH1, DNSRESEARCH2] shows that blocking of fragments is common practice on the Internet today and changing this is a major effort and may take many years if it is achievable at all.

The recommendations in this document are aimed at operators of authoritative name servers. They are encouraged to use the information provided here in configuration tools, or as default values. The recommendations apply until recommendations for resolving name servers are standardised (e.g. [RFC 2671BIS (DRAFT)], implemented and deployed.

Other recommendations regarding the DNS response size exist [RFC 2671, RFC 2671BIS (DRAFT), RFC 3226], which are not obsoleted by this document, but which focus on the behaviour of resolving name servers.

3. Recommended server configuration

1. At least 50% of all authoritative name servers for a zone SHOULD be set to limit the overall response size to 1472 bytes, but MAY be set as low as 1232 bytes;

2. At least 50% of all in-zone authoritative name servers for a zone SHOULD be set to limit the overall response size to 1472 bytes, but MAY be set as low as 1232 bytes;

3. Authoritative name servers to which the above recommendations are applied MUST accept DNS queries over TCP.

[1] Research results in [DNSRESEARCH1] and [DNSRESEARCH2] show that an estimated 2% to 10% of all resolvers experience problems receiving fragmented responses.

4. Remarks and explanation

1. Limiting the response size to 1472 bytes results in responses that can be transported without experiencing fragmentation over paths with an MTU equal to that of Ethernet, or larger;
2. An authoritative name server that is authoritative for a zone is *in-zone* if it has a name in that zone (e.g. ns1.example.com is considered to be in-zone for example.com);
3. Applying these recommendations to too few authoritative name servers for a zone may be insufficiently beneficial for all resolvers, due to various resolution strategies in DNS software;
4. Limiting the response size to 1472 bytes does not affect the functionality of DNS. The reduction in response size is obtained by limiting the number of surplus Additional Records in the responses.

5. Operational considerations

1. Limiting the response size at an authoritative name server can result in a minor increase in truncated UDP responses. Such responses lead to TCP fall-backs. Zones with large RR sets may experience a more significant increase in truncated responses when limiting the response size. Too many TCP fall-backs may negatively affect the availability of an authoritative name server;
2. Reducing the maximum response size to a value below 1472 may result in an undesirable increase in TCP fall-backs [DNSRESEARCH3]. In rare cases a zone may become unreachable when the maximum response size is reduced to 1232 bytes, or below;
3. An operator of an authoritative name server should monitor the increase in TCP fall-backs and how it affects system resources;
4. These recommendations will not directly influence the security of any system with a name in a served zone, or any other system on the Internet. Following these recommendations will generally contribute to reachability of an authoritative name server and hence all systems with a name in a served zone.

6. Example zone configuration

```
example.com. IN NS ns1.example.com.
example.com. IN NS ns2.example.com.
example.com. IN NS ns1.example.net.
example.com. IN NS ns1.example.org.
```

This example zone has 4 authoritative name servers. In order implement the recommendations in this document, the operator must apply these recommendations to ns1.example.com and ns1.example.net. This involves 50% of all authoritative name servers and 50% of all in-zone authoritative name servers for this zone.

7. Acknowledgements

This work is a product of the RIPE DNS Working Group.

8. References

[DNSRESEARCH1] Weaver, N., Kreibich, C., Nechaev, B., and Paxson, V., “Implications of Netalyzr’s DNS Measurements”, Proceedings of the First Workshop on Securing and Trusting Internet Names (SATIN), Teddington, United Kingdom, 2011

[DNSRESEARCH2] Van den Broek, J., Van Rijswijk, R., Pras, A., Sperotto, A., “DNSSEC and firewalls - Deployment problems and solutions”, Private Communication, Pending Publication, 2012

[DNSRESEARCH3] Rikitake, K., Nogawa, H., Tanaka, T., Nakao, K. and Shimojo, S. “An Analysis of DNSSEC Transport Overhead Increase”, IPSJ SIG Technical Reports 2005-CSEC-28, Vol. 2005, No. 33, pp. 345-350, ISSN 0919-6072, 2005

[KAMINSKY1] Musashi, Y., Kumagai, M., Kubota, S. and Sugitani, K., “Detection of Kaminsky DNS Cache Poisoning Attacks”, Proceedings of the 2011 4th International Conference on Intelligent Networks and Intelligent Systems, p. 121-124, Washington, DC, 2011

[NIST1] Scarfone, K., Hoffman, P., “Guidelines on Firewalls and Firewall Policy”, Recommendations of the National Institute of Standards and Technology (NIST), Gaithersburg, MD, 2009

[RFC 2119] Bradner, S., “Key words for use in RFCs to Indicate Requirement Levels”, RFC 2119, March 1997

[RFC 2460] Deering, S., Hinden, R., “Internet Protocol, Version (IPv6) Specification”, RFC 2460, December 1998

[RFC 2606] Eastlake, D., Panitz, A., "Reserved Top Level DNS Names", RFC 2606, BCP 32, June 1999

[RFC 2671] Vixie, P., "Extension Mechanisms for DNS (EDNS0)", RFC 2671, August 1999

[RFC 2671BIS (DRAFT)] Damas, J., Graff, M., Vixie, P., “Extension Mechanisms for DNS (EDNS(0)) draft-ietf-dnsext-rfc2671bis-edns0-09”, RFC 2671bis (DRAFT), August 2012

[RFC 3226] Gudmundsson, O., “DNSSEC and IPv6 A6 aware server/resolver message size requirements”, RFC 3226, December 2001

9. Author's address

Gijs van den Broek
SURFnet bv
PO box 19035
3501 DA Utrecht
Telephone: +31 302 305 305
<gijs.vandenbroek@surfnet.nl>

Roland van Rijswijk
SURFnet bv
PO box 19035
3501 DA Utrecht
Telephone: +31 302 305 305
<roland.vanrijswijk@surfnet.nl>