ENUM   L. Conroy
Internet-Draft   RMRL
Expires: December 28, 2006   J. Reid
                                                                DNS-MODA
                                                           June 26, 2006

ENUM Requirement for EDNS0 Support
<draft-conroy-enum-edns0-02.txt>

Status of this Memo

By submitting this Internet-Draft, each author represents that any applicable patent or other IPR claims of which he or she is aware have been or will be disclosed, and any of which he or she becomes aware will be disclosed, in accordance with Section 6 of BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt.

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.

This Internet-Draft will expire on December 28, 2006.

Copyright Notice

Copyright (C) The Internet Society (2006).

Abstract

Support for EDNS0 (Extension Mechanisms for DNS) is mandated in this document for DNS entities querying for or serving NAPTR records. In general those entities will be supporting ENUM resolution. This requirement is needed because DNS responses to ENUM-related queries generally return large RRSets. Without EDNS0 support these lookups would result in truncated responses and repeated queries over TCP transport. That has a severe impact on DNS server load and on the latency of those queries.

This document updates RFC3761 only in adding this requirement.

Table of Contents

1. Terminology 3
2. Introduction 4
  2.1. DNS - Background 4
3. Problem 6
4. Solution 7
  4.1. Required Aspects of EDNS0 Support 7
    4.1.1. TCP Requirement 8
    4.1.2. Fragmentation Requirement 9
    4.1.3. Intermediary Node Requirement 9
5. Security Considerations 10
6. IANA Considerations 11
7. Acknowledgements 12
8. References 13
  8.1. Normative References 13
  8.2. Informative References 13
Authors' Addresses 15
Intellectual Property and Copyright Statements 16

1. Terminology

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 BCP 14, RFC2119 [12].

2. Introduction

ENUM is defined in RFC3761[1]. It uses the underlying DNS protocol to handle its queries and responses for NAPTR resource records (defined in [2]) that are to be processed by an ENUM client.

The DNS protocol is defined in RFC1034[3], RFC1035[4] and clarified in RFC2181[5]. Requirements for Internet Hosts are specified in RFC1123[6]. DNS is a simple and efficient protocol and is fundamental to the operation of Internet communications.

Entities involved in processing ENUM queries and responses have to deal with messages that typically return large sets of resource records (RRSets). These messages do not fit the profile for which DNS was originally designed, and so it is necessary to implement the standard Extension Mechanisms for DNS as described in RFC2671[7], specifically the feature by which a DNS entity can indicate its ability to process messages of a given size over UDP transport.

2.1. DNS - Background

For historical reasons a size limit of 512 bytes is specified in RFC1035 for all messages exchanged in DNS over UDP transport. Small MTUs were common in early networks and fragmentation issues were not well addressed in the communications software that existed when the DNS was introduced. This 512 byte limit was chosen to avoid the risk of packet fragmentation over paths with a small MTU. When an answer will not fit within this limit, the DNS response will be truncated (indicated by the "TC" flag being set to '1' in the response).

DNS queries and responses can also be carried over TCP transport. In this case, the size limit is not applied because TCP already has robust mechanisms for handling fragmentation and the reconstruction of packets. TCP does have performance implications for simple query-response interactions: for instance by increasing the overall time taken to complete the transaction and increasing the volume of network traffic. Thus it is not the default choice of transport for the DNS protocol.

RFC1035 mandates support for UDP-based queries but only recommends support for TCP-based queries. However RFC1123 essentially makes TCP query support mandatory. It mandates that a DNS resolver discard a truncated response and retry using another transport protocol. In effect, authoritative name servers that do not answer TCP queries after returning truncated responses are misconfigured.

With the introduction of the Extension mechanisms described in RFC2671, there is now a mechanism by which a DNS entity can indicate

that it is capable of handling UDP-based DNS transactions larger than those described in RFC1035.

3. Problem

ENUM zones typically store large sets of resource records (RRSets). An entry for one E.164 telephone number (i.e. one owner-name) could contain 10-20 NAPTR records or more. An answer returning such an RRSet will almost certainly exceed the capacity of a DNS response meeting the size limit set in RFC1035 for messages using UDP transport. RFC 1035 (and RFC1123) outline a "fallback" mechanism. The server indicates that it cannot return the full answer by setting the TC flag in its response. On receiving this message the client will discard the partial result and retry the query using TCP transport.

This fallback induces extra latency and network traffic when resolving ENUM queries. The initial truncated response is returned over UDP and discarded, a TCP transport connection is initiated, the query is repeated and the TCP connection torn down once the complete answer has been received. These overheads are unacceptable in some environments where ENUM will be used: high-latency mobile data networks for instance.

This behaviour also causes extra load on the name servers. They have to process the initial query and construct a truncated response, only to receive the query again using TCP transport. Furthermore, even after it has returned the full answer over a TCP connection the name server must maintain a TCP control block for a certain time after it has sent the answer and shutdown of the TCP connection has been initiated. Answering a high volume of queries using short-lived TCP connections causes issues with memory usage, involves the name server in unnecessary processing and may constrict the number of concurrent connections that may be open. On busy name servers this has severe operational impact on throughput.

The proportion of conventional DNS queries that exceed the UDP size limit specified in RFC1035 is relatively small. So the impact on normal query resolution of this TCP fallback mechanism is minimal. It just does not happen often enough to be a significant concern. However for ENUM lookups for NAPTR records this assumption no longer holds. This fallback procedure will no longer be the exception. It may well be the norm and performance when handling ENUM queries will suffer as a result.

4. Solution

In short, the solution to the problem of returning the large RRSets typical of ENUM queries is to use EDNS0. This will maintain high performance and avoid excessive load on DNS servers. An ENUM client and any resolving name server can use EDNS0 to indicate the size of UDP packet it is prepared to handle in a DNS response. This allows name servers involved in the resolution to return answers using UDP that fit within the limit set by the resolver rather than that specified in RFC1035. For a description of other situations in which EDNS0 is useful and for further motivations on its use, see RFC3225[8] and RFC3226[13].

As well as using EDNS0, it is necessary to ensure that the buffer sizes reported are adequate. It should be noted that the penalty of choosing too low a size for EDNS0 support may be even more severe that the standard method described in RFC 1035 and RFC 1123. Thus it is good practice to select a larger size than is likely to be needed, to counteract that greater cost where fallbacks still occur. Sections 2.4 and particularly 2.5 of [9] explain the rationale for using the size option of EDNS0 for queries that return larger responses. In that document, section 3.1 describes expected server behaviour, section 4.1 describes expected resolver behaviour, while section 3 summarises the proposed message sizes to be supported by servers and resolvers. These same size recommendations are repeated here, as it is felt that ENUM already has a similar issue with larger responses, and will certainly need the larger messages sizes with the introduction of IPv6 and DNSSEC support.

4.1. Required Aspects of EDNS0 Support

There are some subtleties with EDNS0 support within ENUM, so the full implications of the requirement of EDNS0 support for ENUM resolution are explained here.

The basic requirement for EDNS0 support in ENUM entities is in two parts:

> ALL entities involved in querying for or serving NAPTR records MUST support EDNS0.
>
> ALL entities involved in querying for or serving NAPTR records MUST be able to support EDNS0 buffer sizes for queries or responses of at least 1220 bytes, and SHOULD be able to support buffer sizes of 4000 bytes.
>
> Entities querying for NAPTR records MUST use EDNS0 in their queries unless they have current knowledge that EDNS0 support is

> not provided at the target of their queries.
>
> Entities looking up NAPTR records MUST advertise a buffer size of at least 1220 bytes in their queries, and SHOULD advertise a buffer size of 4000 bytes. Consideration should also be given to the MTU of the underlying network, less any overhead needed for lower-level network protocols.

Of course, support is one thing, but use is another. The mandate for support of EDNS0 when processing ENUM queries does not imply spontaneous use. The mechanism described in RFC2671 applies. If a name server receives a query indicating that the client supports EDNS0, then it replies with an extended response, assuming that name server supports ENDS0. If it does not receive such an indication, then it responds with a conventional RFC1035-style reply. Similarly, resolvers querying for NAPTR records must indicate their ability to support EDNS0 and larger buffer sizes when they send those lookups because this is the only way that they will receive such responses.

There are three further aspects to EDNS0 support.

4.1.1. TCP Requirement

Firstly, it is still possible that ENUM-related queries could result in truncated responses and TCP retries even though an EDNS0-enabled mechanism is used. A zone could include a larger set of NAPTR records than will fit into the packet size the client has reported itself as supporting. If the ENUM client requests all available resource records for some ENUM zone rather than just its NAPTR records, there may be large amounts of data for other resource record types for the queried owner-name: eg TXT records. In this case the complete answer may well exceed the client's advertised packet size even though a NAPTR-specific query would not. Also, the EDNS0 query may fail for the reasons covered below. In all these cases the fallback mechanism described in RFC2671 will be needed. For the fallback process to work for large RRSets, entities will need to support TCP transport even if EDNS0 is disabled or unavailable for some reason.

Thus:

> If an entity involved knows that EDNS0 queries and responses work in the current ENUM resolution chain, it MUST be willing to support queries and responses using TCP transport.

4.1.2. Fragmentation Requirement

Second, a DNS server may receive queries that indicate a given size of response is acceptable. However, the resolver may be connected via a network with a lower MTU, in which case the response packet will undergo fragmentation and reassembly in transit.

Thus, although obvious (and not directly related to its use in processing ENUM requests), this means that:

> A DNS server responding to a query that includes the EDNS0 size option MUST NOT set the DF (Don't Fragment) bit in the UDP packet holding its answer.

4.1.3. Intermediary Node Requirement

The final point concerns intermediate nodes. It has been noticed that some intermediate nodes exhibit overly aggressive behaviour.

Specifically:

> Intermediate nodes MUST NOT block or discard valid ENUM queries and responses that indicate EDNS0 support. In particular, intermediate packet filters MUST NOT assume that UDP DNS responses larger than 512 bytes are invalid. These responses are correct and MUST NOT be intercepted provided they comply with the EDNS0 standard. Such packet discard strategies are in error.
>
> Intermediate nodes MUST NOT block valid DNS queries and responses sent over TCP transport. It is perfectly reasonable for DNS queries to be sent over TCP transport.

This last requirement means that intermediary packet filters MUST NOT simply block all TCP-based DNS traffic.

5. Security Considerations

This document does appear to introduce any extra security issues over and above those mentioned in RFC3761 and in RFC2671, as well as those listed in the thorough analysis of the threats to DNS in RFC3833 [14].

It should be noted that mandating the use of EDNS0 by ENUM-related entities also facilitates the deployment of Secure DNS, DNSSEC, currently defined in RFC4033 [10], RFC4034 [11] and RFC4035 [9]. Secure DNS will be necessary to verify the integrity of ENUM responses. RFC3225 [8] states that clients signal their ability to handle signed responses via the DO (DNSSEC OK) bit in the EDNS0 header and a name server will not return these unless this bit is set. So unless EDNS0 is used, ENUM-related entities will be unable to verify DNSSEC-signed responses from the DNS. Signed replies from the DNS are also much larger than unsigned ones, which provided an added incentive to use larger UDP payloads.

6. IANA Considerations

This document has no IANA requirements.

7. Acknowledgements

We would like to thank the working group members active on the ENUM mailing list who engaged in this topic, and the development and operational teams that collected data confirming the need for this mandate.

8. References

8.1. Normative References

[1] Faltstrom, P. and M. Mealling, "The E.164 to Uniform Resource Identifiers (URI) Dynamic Delegation Discovery System (DDDS) Application (ENUM)", RFC 3761, April 2004.

[2] Mealling, M., "Dynamic Delegation Discovery System (DDDS) Part Three: The Domain Name System (DNS) Database", RFC 3403, October 2002.

[3] Mockapetris, P., "DOMAIN NAMES - CONCEPTS AND FACILITIES", RFC 1034, November 1987.

[4] Mockapetris, P., "Domain names - implementation and specification", STD 13, RFC 1035, November 1987.

[5] Elz, R. and R. Bush, "Clarifications to the DNS Specification", RFC 2181, July 1997.

[6] Braden, R., "Requirements for Internet Hosts -- Application and Support", RFC 1123, October 1989.

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

[8] Conrad, D., "Indicating Resolver Support of DNSSEC", RFC 3225, December 2001.

[9] Arends, R., Austein, R., Larson, M., Massey, D., and S. Rose, "Protocol Modifications for the DNS Security Extensions", RFC 4035, March 2005.

[10] Arends, R., Austein, R., Larson, M., Massey, D., and S. Rose, "DNS Security Introduction and Requirements", RFC 4033, March 2005.

[11] Arends, R., Austein, R., Larson, M., Massey, D., and S. Rose, "Resource Records for the DNS Security Extensions", RFC 4034, March 2005.

8.2. Informative References

[12] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", RFC 2119, BCP 14, March 1997.

[13] Gudmundsson, O., "DNSSEC and IPv6 A6 Requirements", RFC 3226,

December 2001.

[14] Atkins, D. and R. Austein, "Threat Analysis of the Domain Name System (DNS)", RFC 3833, August 2004.

[15] Bradner, S., "The Internet Standards Process -- Revision 3", RFC 2026, BCP 9, October 1996.

[16] Bradner, S., "IETF Rights in Contributions", BCP 78, RFC 3978, March 2005.

[17] Bradner, S., "Intellectual Property Rights in IETF Technology", BCP 79, RFC 3979, March 2005.

Authors' Addresses

Lawrence Conroy
Roke Manor Research
Roke Manor
Old Salisbury Lane
Romsey
United Kingdom

Phone: +44 1794 833666
Email: lconroy@insensate.co.uk
URI: http://www.sienum.co.uk

Jim Reid
DNS-MODA
DNS-MODA
6 Langside Court
Bothwell, SCOTLAND
United Kingdom

Phone: +44 1698 852881
Email: jim@dns-moda.org

Intellectual Property Statement

The IETF takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights. Information on the procedures with respect to rights in RFC documents can be found in BCP 78 and BCP 79.

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at http://www.ietf.org/ipr.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.

Disclaimer of Validity

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Copyright Statement

Copyright (C) The Internet Society (2006). This document is subject to the rights, licenses and restrictions contained in BCP 78, and except as set forth therein, the authors retain all their rights.

Acknowledgment

Funding for the RFC Editor function is currently provided by the Internet Society.