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 April 25, 2006.

Copyright Notice

Copyright (C) The Internet Society (2005).

Abstract

This document mandates support for EDNS0 (Extension Mechanisms for DNS) in DNS entities claiming to support ENUM query resolution (as defined in RFC3761). This requirement is needed as DNS responses to ENUM-related questions return larger sets of Resource Records than typical DNS messages. Without EDNS0 support in all the involved entities, a fallback to TCP transport for ENUM queries and responses would typically occur. That has a severe impact on DNS Server load,
and on latency of ENUM 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 ..................... 8
      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 [8].
2. Introduction

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

The DNS protocol is defined in RFC1034[3], RFC1035[4] and clarified in RFC2181[5], whilst Requirements for Internet Hosts are specified in RFC1123[6]. It 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

DNS is based on a simple question and answer model. In the standard approach described in RFC1035, a Resolver will construct and send a question to a DNS Server using UDP transport.

For historical reasons, a size limit of 512 bytes is specified in RFC1035 for all messages exchanged in DNS over UDP transport, originally to avoid the risk of packet fragmentation over paths with a small MTU, as handling such fragmentation was not well supported in early communications software. Where an answer will not fit within this limit, a basic 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, as TCP already has a robust mechanism for fragmentation and reconstruction of packets, including the selective re-transmission of fragments that have been corrupted or discarded in transit. It does have performance implications, 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, but can be used if a question sent using UDP fails due to truncation of the response.

Supporting UDP queries is mandatory, but support for TCP queries is recommended also, and is (in effect) required as RFC1123 requires 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 messages larger than those implied in the scheme described in RFC1035, so that it can use UDP transport but still receive DNS messages up to the size it specifies in its request or response.
3. Problem

ENUM zones typically store large sets of Resource Records (RRSets), and an answer returning such an RRSets may well exceed the capacity of a DNS response meeting the size limit set in RFC1035 for messages using UDP transport. RFC1035 (and RFC1123) outline a fallback mechanism; the Server indicates that it cannot return the full answer by setting the TC flag in its response, and, on receiving this message, the Resolver will discard the partial result and use TCP transport when re-trying its question.

This fallback does induce extra latency in resolving the question, as the Resolver waits for the truncated response and then opens a TCP connection over which to re-try its question. It generates extra data traffic, as the initial truncated response is returned and immediately discarded, the TCP transport connection creation and release are both signalled, and the query is sent twice before the final answer is returned.

It also causes extra load on the DNS Server. The DNS Server has to process the initial query and construct a truncated response, only to receive the query again using TCP transport. Likewise, even after it has returned the full answer over a TCP connection opened by the Resolver, the Server must maintain a TCP control block for a certain time after it has sent the answer and has requested closure of the TCP connection. Thus answering a high volume of queries using TCP connections causes issues with memory usage, involves the Server in unnecessary processing and, due to the linger time required before each connection can be released, limits the number of concurrent connections that may be open.

While the percentage of queries processed that exceed the UDP size limit specified in RFC1035 is relatively small, the impact on normal query resolution of the fallback mechanism is minimal; it just does not happen that often. However, where ENUM queries are concerned, this design assumption no longer holds; this fallback procedure is no longer the exception, and performance when handling ENUM queries suffers as a result.
4. Solution

In short, ENUM queries return responses containing large RRsets. If ENUM clients use the query mechanism outlined in RFC1035, it will create significant levels of truncated responses and query retries over TCP. These can have a serious operational impact: poor performance for clients and increased load on DNS servers. The solution to this problem is for all entities that are involved in ENUM resolution to use EDNS0. The ENUM client (and the DNS Resolver it uses) will indicate its ability to accept large responses by adding to its query an OPT pseudo-resource record as additional data, showing the size of UDP packet it can process in response. This allows the DNS Servers involved in the resolution to return answers that fit within the limit set by the Resolver rather than that specified in RFC1035, whilst still using UDP as the transport protocol. For a description of other situations in which EDNS0 is useful and for further motivations on its use, see RFC3225[9] and RFC3226[10].

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 ENUM resolution MUST support EDNS0

Resolvers MUST use EDNS0 in their queries unless they have current knowledge that EDNS0 support is not provided at the target of their queries.

Note that ENUM-involved entities includes ALL the DNS Servers and Resolvers used in ENUM resolution. The Resolvers involved include both the Stub Resolver typically installed on a client node, plus the Recursive Resolver to which it forwards requests and from which it receives answers. The list of involved Servers includes any DNS Servers authoritative for delegated ENUM domains, but also the DNS Servers authoritative for .arpa, the ENUM Tier 0 DNS Servers authoritative for .e164.arpa, and all ENUM Tier 1 DNS Servers with zones delegated from those Tier 0 Servers.

Of course, support is one thing, but use is another. To clarify, the mandate for Server support of EDNS0 when processing ENUM queries does not imply spontaneous use. Such a Server operates in exactly the way described in RFC2671. A DNS Server that is involved in ENUM
resolution assumes nothing about Resolver support for EDNS0; the Resolver will indicate this in its query. If a Server receives a query indicating that the Resolver supports extended responses, then it replies with an extended response; if it does not receive such an indication, then it responds with a conventional RFC1035-style reply.

Equally, Resolvers involved in ENUM resolution MUST indicate their ability to process extended responses when they send ENUM queries - this is the only way that they will receive such responses.

There are three further aspects to EDNS0 support.

4.1.1. TCP Requirement

First, even with an EDNS0-enabled mechanism, it is still possible that a query will not succeed. It is possible for a zone to include a larger set of NAPTRs than will fit into the packet size the Resolver has reported itself as supporting. Similarly, the ENUM client may have requested all available resource records, rather than just the NAPTRs. If for example the queried zone contains several TXT records, the complete answer may exceed the reported packet size supported even though a NAPTR-specific query would succeed. Also, the EDNS0 query may fail for the reasons covered next. In all these cases the fallback mechanism described in RFC2671 will be needed.

For that fallback process to work for large RRSets, entities will need to support TCP transport even if EDNS0 is disabled for some reason.

Thus:

Unless an entity involved knows that EDNS0 queries and responses work in the current ENUM resolution chain, it SHOULD 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 valid ENUM queries and responses that indicate EDNS0 support, even if these are larger than the size given in RFC1035. In particular, intermediate packet filters MUST NOT assume that large DNS queries and responses are invalid; they are not if they indicate EDNS0 support correctly. Such packet discard strategies are in error.

Intermediate nodes MUST NOT block valid ENUM queries and responses sent over TCP transport.

This last requirement means that intermediary packet filters MUST NOT simply block all TCP traffic; it is perfectly reasonable for DNS queries to be sent over TCP transport, and a more selective strategy will need to be chosen.
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 [11].

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 [12], RFC4034 [13] and RFC4035 [14]. Secure DNS will be necessary to verify the integrity of ENUM responses. RFC3225 [9] 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 provides 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, the development and operational teams that collected data confirming the need for this mandate, and Alexander Mayrhofer for his detailed review and helpful comments.
8. References

8.1. Normative References


8.2. Informative References


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 (2005). 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.