Network Working Group
Internet Draft
Expiration Date: November 1996

Robert Elz
University of Melbourne

Randy Bush
RGnet, Inc.

May 1996

Clarifications to the DNS Specification

draft-ietf-dnsind-clarify-01.txt

Status of this Memo

1. Abstract

This draft considers some areas that have been identified as problems with the specification of the Domain Name System, and proposes remedies for the defects identified. Two separate issues are considered, IP packet header address usage from multi-homed servers, and TTLs in sets of records with the same name, class, and type.

2. Introduction

Several problem areas in the Domain Name System specification [RFC1034, RFC1035] have been noted through the years [RFC1123]. This draft addresses two additional problem areas. The two issues here are independent. Those issues are the question of which source address a multi-homed DNS server should use when replying to a query, and the issue of differing TTLs for DNS records with the same label, class and type.

Suggestions for clarifications to the DNS specification to avoid these problems are made in this memo. The solutions proposed herein are intended to stimulate discussion. It is possible that the sense of either may be reversed before the next iteration of this draft, but less likely now than it was before the previous version.

3. Server Reply Source Address Selection

Most, if not all, DNS clients, whether servers acting as clients for the purposes of recursive query resolution, or resolvers, expect the address from which a reply is received to be the same address as that to which the query eliciting the reply was sent. This, along with the identifier (ID) in the reply is used for disambiguating replies, and filtering spurious responses. This may, or may not, have been intended when the DNS was designed, but is now a fact of life.

Some multi-homed hosts running DNS servers fail to anticipate this usage, and consequently send replies from the "wrong" source address, causing the reply to be discarded by the client.

3.1. UDP Source Address Selection

To avoid these problems, servers when responding to queries using UDP must cause the reply to be sent with the source address field in the IP header set to the address that was in the destination address field of the IP header of the packet containing the query causing the response. If this would cause the response to be sent from an IP address which is not permitted for this purpose, then the response may be sent from any legal IP address allocated to the server. That address should be chosen to maximise the possibility that the client will be able to use it for further queries. Servers configured in such a way that not all their addresses are equally reachable from all potential clients need take particular care when responding to queries sent to anycast, multicast, or similar, addresses.

3.2. Port Number Selection

Replies to all queries must be directed to the port from which they were sent. With queries received via TCP this is an inherent part of the transport protocol, for queries received by UDP the server must take note of the source port and use that as the destination port in the response. Replies should always be sent from the port to which they were directed. Except in extraordinary circumstances, this will be the well known port assigned for DNS queries [RFC1700].

4. Resource Record Sets

Each DNS Resource Record (RR) each has a label, class, type, and data. While it is meaningless for two records to ever have label, class, type and data all equal (servers should suppress such duplicates if encountered), it is possible for many record types to exist with the same label class and type, but with different data. Such a group of records is hereby defined to be a Resource Record Set (RRSet).

4.1. Sending RRs from an RRSet

A query for a specific (or non-specific) label, class, and type, will always return all records in the associated RRSet - whether that be one or more RRs, or the response shall be marked as "truncated" if the entire RRSet will not fit in the response.

4.2. TTLs of RRs in an RRSet

Resource Records also have a time to live (TTL). It is possible for the RRs in an RRSet to have different TTLs, however no uses for this have been found which cannot be better accomplished in other ways. This can, however, cause partial replies (not marked "truncated") from a caching server, where the TTLs for some but not all of the RRs in the RRSet have expired.

Consequently the use of differing TTLs in an RRSet is hereby deprecated, the TTLs of all RRs in an RRSet must be the same.

Should a client receive a response containing RRs from an RRSet with differing TTLs, it should treat the RRs for all purposes as if all TTLs in the RRSet had been set to the value of the lowest TTL in the RRSet.

4.3. Receiving RRSets

Servers never merge RRs from a response with RRs in their cache to form an RRSet. If a response contains data which would form an RRSet with data in a server's cache the server must either ignore the RRs in the response, or use those to replace the existing RRSet in the cache, as appropriate. Consequently the issue of TTLs varying between the cache and a response does not cause concern, one will be ignored.

4.3.1. Ranking data

When considering whether to accept an RRSet in a reply, or retain an RRSet already in its cache instead, a server should consider the relative likely trustworthiness of the various data. That is, an authoritative answer from a reply should replace cached data that had been obtained from additional information in an earlier reply, but additional information from a reply will be ignored if the cache contains data from an authoritative answer or a zone file.

The accuracy of data available is assumed from its source. Trustworthiness shall be, in order from most to least:

+ Data from a primary zone file, other than glue data,
+ Data from a zone transfer, other than glue,
+ That from the answer section of an authoritative reply,
+ Glue from a primary zone, or glue from a zone transfer,
+ Data from the authority section of an authoritative answer,
+ Data from the answer section of a non-authoritative answer,
+ Additional information from an authoritative answer,
+ Data from the authority section of a non-authoritative answer,
+ Additional information from non-authoritative answers.

Where authenticated data has been received it shall be considered more trustworthy than unauthenticated data of the same type.

"Glue" above includes any record in a zone file that is not properly part of that zone, including nameserver records of delegated sub-zones (NS records), address records that accompany those NS records (A, AAAA, etc), and any other stray data that might appear.

4.4. Sending RRSets (reprise)

A Resource Record Set should only be included once in any DNS reply. It may occur in any of the Answer, Authority, or Additional Information sections, as required, however should not be repeated in the same, or any other, section, except where explicitly required by a specification. For example, an AXFR response requires the SOA

record (always an RRSet containing a single RR) be both the first and last record of the reply. Where duplicates are required this way, the TTL transmitted in each case must be the same.

5. Security Considerations

This document does not consider security.

In particular, nothing in section 3 is any way related to, or useful for, any security related purposes.

Section 4.3.1 is also not related to security. Security of DNS data will be obtained by the Secure DNS [DNSSEC], which is orthogonal to this memo.

It is not believed that anything in this document adds to any security issues that may exist with the DNS, nor does it do anything to lessen them.

6. References

[RFC1034] Domain Names - Concepts and Facilities, (STD 13)
P. Mockapetris, ISI, November 1987.

[RFC1035] Domain Names - Implementation and Specification (STD 13)
P. Mockapetris, ISI, November 1987

[RFC1123] Requirements for Internet hosts - application and support, (STD 3) R. Braden, January 1989

[RFC1700] Assigned Numbers (STD 2)
J. Reynolds, J. Postel, October 1994.

[DNSSEC] Domain Name System Security Extensions,
D. E. Eastlake, 3rd, C. W. Kaufman,
Work in Progress (Internet Draft), January 1996.

7. Acknowledgements

This memo arose from discussions in the DNSIND working group of the IETF in 1995 and 1996, the members of that working group are largely responsible for the ideas captured herein.

8. Authors' addresses

Robert Elz
Computer Science
University of Melbourne
Parkville, Victoria, 3052
Australia.

EMail: kre@munnari.OZ.AU

Randy Bush
RGnet, Inc.
9501 SW Westhaven
Portland, Oregon, 97225
United States.

EMail: randy@psg.com