DNSOP Working Group
Internet Draft
Expires: January 2015

Xiaodong Lee
CNNIC
Vixie Paul
Farsight Security, Inc.
Zhiwei Yan(Ed.)
CNNIC

July 3, 2014

How to scale the DNS root system?
draft-lee-dnsop-scalingroot-00.txt

Abstract

In Domain Name System (DNS), 13 root servers are deployed in order to provide the entrance of domain name resolution and they are denoted by 13 letters. From 2000 or so, the anycast technology has been adopted to extend the number of root servers with the explosive development of Internet. To date, 11 of the 13 letters are hosted at multiple sites, and the root zone is served at about 380 sites around the globe. However, increasing mirror sites is not a perfect solution to scale the DNS root servers because the geographical distribution of the current 13 root servers is uneven and only increasing mirror sites cannot solve the efficiency and scalability issues caused by that. Then we propose a new DNS root system in this draft based on the widely deployed Domain Name System Security Extensions (DNSSEC). The proposed architecture is scalable and secure enough to meet the current and future needs to scale the DNS root system in an easy-deployment way.

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 [RFC2119].

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and 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

Copyright Notice

Table of Contents

1. Introduction

In Domain Name System (DNS), 13 root servers are deployed and they are denoted by 13 letters from A to M. From 2000 or so, the anycast technology has been adopted to extend the number of root servers with the explosive development of Internet. To date, 11 of the 13 letters are hosted at multiple sites, and the root zone is served at about 380 sites around the globe [1,2].

With the continuous development of Internet, more and more DNS root servers have to be deployment in order to meet the efficiency and stability requirements of the DNS root system. On one hand, the DNS resolution efficiency should be high enough to support more and more Internet users and their sophisticated applications. On the other hand, the DNS root system has to be stable enough to face the cyber attacks which are becoming more and more serious [3]. Then, how about deploying more anycast mirror sites? Without doubt, this scheme can improve the efficiency and stability of DNS root system as usual, but its problems are:

a) This scheme is not open enough to satisfy the special requirements of a country or district or organization (CDO). If a CDO wants to deploy a root server to serve its local users, the current procedure is complex and it is almost impossible for the CDO to control the deployed site no matter the type of the site is local or global. Specially, the current root name servers have respective Root Name Server Operators (RNSOs), to whom IP address blocks have been allocated. The CDO wants a root server mirror site must coordinate anycast routing configuration and server placement with the corresponding RNSO.

b) This scheme is not stable enough. When one root server is attacked and fails down, its anycast mirrors cannot do the zone file synchronization and the DNS root service may be affected.

2. Motivations

During recent years, many CDOs declare their anticipation to host a DNS root server. And how to adjust the placement of the current 13 DNS root servers also drew a lot of attention from both academic and technical worlds [2]. However, it's very difficult to change the current belonging situation of DNS root servers. If possible, it's a good idea to deploy more root servers (not just the mirror sites) to solve or relieve the above problems.

In the IPv4 based Internet, only 13 root servers are allowed due to the 512-byte limitation. However, adding the IPv6 address information for more than two root name servers will increase the size of the DNS priming response to exceed this maximum. Ultimately, when all 13 root name servers assign IPv6 addresses, the priming response will be increased to 811 bytes [3]. Specially, if a message cannot fit in the 512 bytes, the server must set a special flag indicating that the message was truncated. When receiving such a message, a DNS resolver will normally retry the transaction using TCP instead of UDP [4]. However, the costs of using TCP rather than UDP, in terms of system and network resources, are much higher and

can have significant impact on systems such as name servers that may receive several thousands of queries per second during normal operations. At the same time, with the promotion from ICANN, work is in progress to put forward a recommendation that will enable the publication of IPv6 addresses for, at least, DNS root servers in as short a time as possible. Besides, with the exhaustion of IPv4 address, the IPv6 is promoted from all parties in the community. Currently, 10 IPv6 addresses have been configured on the root servers and the total size of the DNS priming response message has been increased to above 512 bytes (We even do not consider the DNSSEC herein, which increases the DNS response message more tremendously).

Above background means that the 512-byte limitation does not actually exist anymore in the DNS [5]. Then how many DNS root server can be increased further? We take the IPv6 supported Internet as an example, then the DNS priming response message structure is shown in the following if we set the maximum number of root servers is N [6].

--Required Header: 12 bytes

--Query: 5 bytes

--Answer: 31+15*(N-1) [the first answer is 31 bytes, the sequent answers are reduced by 16 bytes due to compression]

--Additional Records: 16*N [each A record is 16 bytes]

--Additional Records: 28*N [each AAAA record is 28 bytes]

The IPv6 supported node is not required to reduce the size of subsequent packets to less than 1280 bytes, but must include a "Fragment Header" in those packets so that the IPv6-to-IPv4 translating router can obtain a suitable "Identification" value to use in resulted IPv4 fragments. Note that this means the payload may have to be reduced to 1232 bytes (1280 minus 40 for the IPv6 header and 8 for the Fragment header) [7].

Then we get the following formula,

$$12+5+31+15*(N-1)+16*N+28*N<1232$$

Accordingly, N=20, which means that the root servers can only be increased by 7 more in the IPv6 transport environment (without EDNS0 [8] and TCP supporting).

In advance, DNSSEC should be considered because it is necessary to guarantee the security of the DNS information and it has been widely deployed (and will be more widely deployed in the future). If only one kind of signature algorithm is used to generate the RRSIG resource record, the size of DNS response message will be increased to about 7 to 10 times comparing with the original one without signing [9]. It means that it is unpromising to increase the root servers further in the current architecture.

In the security aspect, criticisms of the current and historical root name server system is lack of resistance to DDoS attack, noting that even with the current wide scale anycasting by every RNSO, there are still only a few hundred name servers in the world to answer authoritatively for the DNS root zone. We are also concerned that reachability of the root name server system is required even for purely local communication, since otherwise local clients have no way to discover local services. In a world sized distributed system like the Internet, critical services such as DNS root system ought to be extremely well distributed.

In a word, we have to design a new architecture to scale the DNS root name server system and in this way the root server deployment balance can be reconsidered (the future deployment can refer the actual requirements such as the number of Internet users, amount of Internet traffic and so on). Besides, this architecture should scale the DNS root servers without the technology limitations as illustrated above.

3. Proposed architecture

The proposed architecture is strongly based on the widely deployed DNSSEC and shown in Figure 1. With DNSSEC, it is now possible for recursive name server operators to configure DNSSEC validation, such that any gTLD information heard from a root server (or mirror site) must be IANA-approved as indicated by DNSSEC signatures made with IANA's root Zone Signing Key (ZSK).

For the universal anycast root server nodes deployment, we here provide two actual solutions:

1)
The DNS root service manager (may still be the IANA) would generate and digitally sign (with DNSSEC) an additional version of the root zone that has a different set of NS records at its apex. These NS records will denote name servers whose addresses are not assigned to any current RNSO but are instead held in trust by IANA for use by any or all interested CDOs (Global Root X in Figure 1). IANA would request infrastructure micro-allocations from an RIR

(such as ARIN or APNIC), as several IPv4 24-bit prefixes and several IPv6 48-bit prefixes, for use in universal anycasting of the root zone. For example, the following configuration can be used corresponding to the universal root server (Global Root X):

```
. IN NS anycast-X1.iana-servers.net.

. IN NS anycast-X2.iana-servers.net.

$ORIGIN iana-servers.net.

anycast-x1 IN AAAA 2001:?:1::1

anycast-x1 IN A ?.?.1.1

anycast-x2 IN AAAA 2001:?:2::2

anycast-x2 IN A ?.?.2.2
```

2) Based on the contract or approval of one or multiple RNSO, the related root server can also be globally anycasted or locally deployed and totally managed and controlled by a CDO (Root A1 in Figure 1). This case is backward-compatible and does not need to change the current DNS root system.

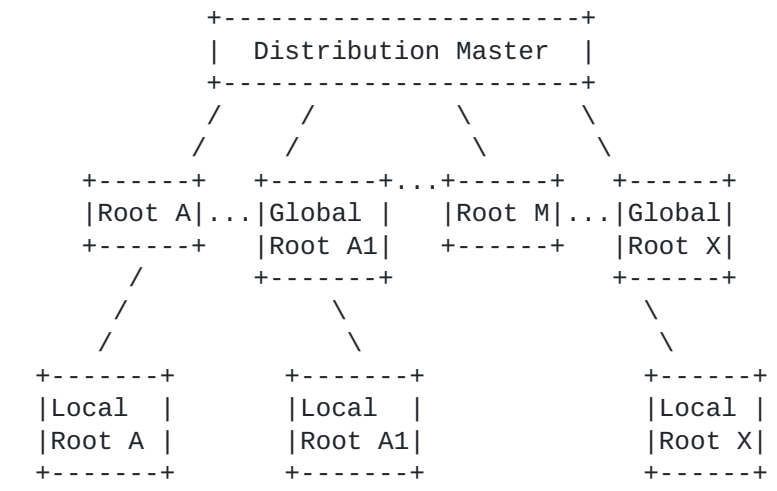

Figure 1. DNS root server architecture

We herein assume that the DNSSEC is widely deployed. In this way, the root zone file will not be tampered or misused. If the DNS root

server is a local site (Local Root A) and only serve for the restricted area, it may be unnecessary for holding the global anycast address. Still in this case, the local site (Local Root A) can also expand its site to multiple mirrors within the local area (for example, within one ISP's coverage) as the global node (Global Root A1) does.

Accordingly, the DNS request message may be served by any root sever during its transmission and the possible scenarios are shown in Figure 2.

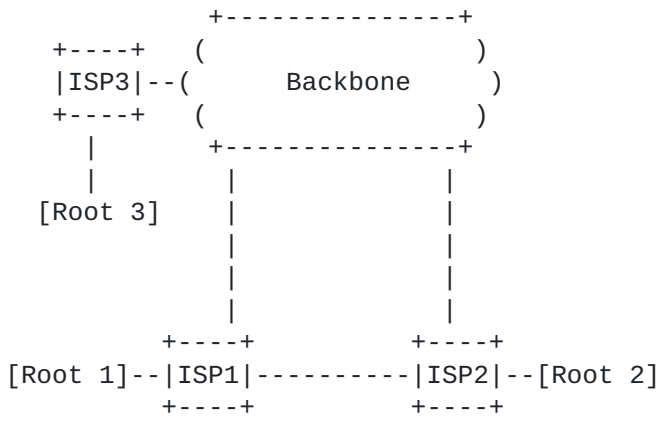

Figure 2. DNS resolution scenarios

We assume that one ISP only holds one DNS root server. In the first case, the DNS request message originated from the edge user is served by the root server deployed in the local network within the same ISP (Root 1) and then the response can be served as soon as possible. In the second case, the DNS request is transmitted to the nearby ISP due to the routing protocol and responded by the root server in the nearby ISP (Root 2). In the third case, the request may travel all the way to the Internet backbone without having been answered closely, in which case it may be answered by an RNSO who has decided to advertise a route to the well known address of the CDO root server (Root 3).

For the data synchronization, the current scheme can be used. After the root zone file is produced, the Distribution Master (DM) (such as the current hidden server) sends a DNS notify message to all root servers and then every root server responds with an acknowledgement to the DM. The DNS notify triggers the Start of Authority (SOA) request. The DM replies with a message which includes the serial number. If this serial number is higher than the number that is

currently operated by the root server, the root server starts the Zone Transfer (XFR) to retrieve the new root zone file. If this serial number in the SOA response is not higher than the root server's currently used version, the root server will take no further action. If this scheme is adopted, each CDO root server has to register with the DM in order to receive the notify message and which is the requirement of the first deployment solution illustrated above. While, for the second deployment solution, the DM function may be the root server managed by the current RNSO.

Of course, a more open scheme is needed in the future for the root zone file synchronization. Because the root zone file data is solid and cannot be tampered, the CDO can actively fetch the root zone file data according to its special requirement and configuration. We mention this synchronization type because the CDO root servers may be increased significantly in the future and the traditional scheme explained above is not flexible and efficient enough.

In a word, DNSSEC makes it possible to deploy the DNS root servers in a more distributed and flat manner, because any server can synchronize the root zone file without modification.

4. Management considerations

Considering the deployment of the architecture proposed in this draft, the following management questions should be answered:

1) How to manage the universal anycast DNS root servers?

We introduce the idea to globally distribute the root servers and locally deploy multiple local nodes. In the routing aspect, the anycast addresses will be broadcasted more widely for the global nodes to be accessed anywhere. However, the addresses of the local nodes have to be controlled in the local area (with the no-export attribute in BGP routing protocol) to serve the requests from the particular local area.

2) What are the principles to deploy the universal DNS root servers?

The principles or rules to select the service provider (CDO) and deploy the root servers can refer to [10,11]. Besides, more actual aspects can be considered in this new architecture due to its minimized limitation to deploy a root server.

Other management and deployment issues will be added further.

5. Security considerations

Although this architecture maintains the basic operations of the current DNS root system and follows the standard DNS protocols, it changes the current DNS root system because we want this mature system to be flexibly scaled with the Internet. Then the following issues related to security and stability may be caused:

1) Routing table: according to this architecture, more than 13 root servers may broadcast their IP addresses globally. This will increase the entries of the BGP routing table (even maybe only one pair of additional anycast addresses (IPv4 and IPv6)).

2) Attack defense: due to the deployment of anycast servers is more widely and the anycast nodes are increased a lot. How to secure these nodes will be harder and more important. In order to manage the increased global/local nodes, more sophisticated tools (such as CHOAS resource record, Traceroute command, special monitoring and analyzing platform) should be adopted and developed. In this way, the anycast nodes can be identified and monitored effectively and efficiently.

Of course, we believe that the future DNS root service provider (many ccTLD and gTLD have deployed anycast service and manage it well) for the global or local anycast nodes will have the experience and ability to monitor and manage the servers and the possible attack (such as DDoS) can be effectively detected and defended [12].

Other security issues will be discussed and detailed further.

6. References

[1] http://www.root-servers.org/.

[2] T. Lee, B. Huffaker, M. Fomenkov, and k. claffy. On the problem of optimization of DNS root servers' placement. In Proc. of Passive and Active Network Measurement Workshop (PAM). April 2003.

[3] P. Vixie and A. Kato. DNS Response Size Issues. draft-ietf-dnsop-respsize-15.txt. February 2014.

[4] R. Bellis. DNS Transport over TCP - Implementation Requirements. IETF RFC 5966. August 2010.

[5] CAIDA. Analysis of the DNS root and gTLD nameserver system: status and progress report. May 2008.

[6] ICANN. Accommodating IP Version 6 Address Resource Records for the Root of the Domain Name System. January 2007.

[7] S. Deering, and R. Hinden. Internet Protocol, Version 6 (IPv6) Specification. IETF RFC 2460. December 1998.

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

[9] http://www.cisco.com/web/about/ac123/ac147/archived_issues/ipj_7-2/dnssec.html.

[10] S. Sato, T. Matsuura, and Y. Morishita. BGP Anycast Node Requirements for Authoritative Name Servers draft-sato-dnsop-anycast-node-requirements-02.txt. September 2007.

[11] R. Bush, D. Karrenberg, M. Kosters, and R. Plzak. Root Name Server Operational Requirements. IETF RFC 2870. June 2000.

[12] USC/ISI Technical Report. Identifying and Characterizing Anycast in the Domain Name System. May 2011.

Author's Address

Xiaodong Lee
China Internet Network Information Center
No.4 South 4th Street, Zhongguancun
Beijing, P. R. China
Email: xl@cnnic.cn

Paul Vixie
Farsight Security, Inc.
155 Bovet Road, #476
San Mateo, CA 94402, USA
Email: vixie@farsightsecurity.com

Zhiwei Yan
China Internet Network Information Center
No.4 South 4th Street, Zhongguancun
Beijing, P. R. China
Email: yanzhiwei@cnnic.cn

Acknowledgment

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