Tokyo Institute of Technology
February 2004

Root Name Servers with Inter Domain Anycast Addresses

Status of this Memo

Abstract

This memo describes an operational guideline for millions of name servers to share an interdomain anycast address.

It enables people operate as many root name servers as they want and still make them traceable.

1. Motivation

DNS root servers are the essential component of the Internet that all the ISPs in the world want to run several root servers.

To satisfy them, we need to have thousands or millions of root servers.

However, because of the restriction on DNS message size over UDP, the number of unicast IP addresses of root servers is severely limited.

Thus, it is necessary to increase the number of root servers by assigning an IP address to a lot of root servers.

Even if DNS were designed to allow a lot of root servers, it is difficult for DNS clients to choose the best (with regard to availability, credibility of zone content, delay, domain policy etc.) root servers among so many root servers. It is not practical to ping millions of servers to find which has the smallest RTT.

This memo proposes a mechanism of policy based selection of a root server sharing an IP address (anycast IP address) with other root servers and discusses operational issues related to it.

Because the selection is policy based, domain administrators have some control over the selection of the best root server among root servers sharing an IP address.

Note that operations similar to that described in this memo are possible today locally without global coordination by any operator who may be irritated by the lack of his control on (sufficiently many) root servers, which may be a source of some operational problems. This memo is an attempt to document the way to solve the problem in a least harmful manner.

Similar operation described in this memo may be applicable to gTLD or other global servers but it is outside the scope of this memo.

2. Suggested Operation

As is demonstrated by proliferated private use addresses, it is easy to set up routers to let unicast addresses have local scopes. It is also easy to let the unicast addresses have nested local scopes. The important difference between the addresses for private use and root servers is in their semantics that the root servers sharing an address also share the globally unique semantics of the address. The root servers may share a globally unique DNS host name, too.

A possible problem of such addresses is that the shared addresses can not be used for global communication.

So, the root name servers with the anycast addresses must have additional globally unique unicast address (or addresses), which may be used for global communication such as zone transfer.

The other possible problem of such addresses is that the shared addresses are not managed by a single entity that the mapping from the shared address of a root server to some operational entity is impossible.

However, if a router adjacent to (or near) the root server has a globally unique address, it is possible to map from the global address to an operational entity, which is expected to be operating the root server. That is, tools like "traceroute" work to uniquely identify the operational entity of the root servers sharing a anycast address.

To be compatible with the current practice that a single address belong to a single AS, each anycast address is assigned its own AS number. There will be multiple ASes of the AS number containing the same address ranges.

ASes, still, can be identified by adjacent ASes. For example, network operators may choose their favorite root server based on the AS numbers of the next hop ASes with, for example, AS path and BGP policy.

It is required that operators of an AS adjacent to the root servers' AS be fully responsible to the operation of the root servers. If a root server's AS is adjacent to multiple ASes, operators of all the ASes must be fully responsible to the operation of the root server. Thus, if there is a routing problem related to a root server, operators of the next hop AS(es) should be contacted.

To avoid complex routing tricks, globally unique unicast address(es) of the root name servers must be taken from the AS(es) adjacent to the root server's AS. Then, in a likely simple case that the root server's AS consists of a single host, which acts as a name server and a BGP router, the host should peer with adjacent AS(es) through an interface(s) address(es) of which belongs to the adjacent AS(es). If the root server's AS has more complex structure, special IGP arrangement of globally unique unicast address(es) is necessary in the AS and at the border router(s) of the adjacent AS(es). The border router(s) must accept IGP information advertised from the root server's AS.

3. Redundancy Considerations

There is widespread misunderstanding on anycast (and multicast) in, including but not limited to, RFC1546 and RFC2461 that anycast (and multicast) could have provided meaningful redundancy or fault tolerance.

It is true that anycast and multicast tolerate some route faults.

However, a fault mode where a server process crash on an anycast server a route to which is still alive, can not be tolerated.

Multicast, at least scalable one, is no better, because scalable multicast needs some multicast server, such as a rendez vous point or a core, which is the single point of failure.

Redundancy with no single point of failure can only be provided by using multiple anycast (or multicast) addresses served by different anycast (or multicast) servers.

Thus, it is meaningless that RFC1546 considers a case where there are multiple anycast servers on a single subnet, because of redundancy. Like unicast, it is a configuration error if there are two or more anycast servers sharing an anycast address in a subnet, which means that anycast works with IPv4 ARP and no special treatment of ND in RFC2461 is necessary.

4. Assignment

As is discussed in the previous section, when a server with an anycast address fails but a route to it is still available, there is no way for clients use other servers with the same anycast address. That is, anycast does not improve availability of servers so much.

So, even with anycast addresses, there should be multiple root servers.

However, as anycast solves the problem of load concentration, we don't need so many anycast IP addresses,

We should have at least 3 and at most 7 anycast addresses for root servers.

5. Security Considerations

This memo describes just an operational guideline with no protocol change. As such, the guideline does not introduce any security issues of the protocol level.

As the route forgery to the root servers can be implemented today without this memo by anyone including local intruders, the guideline does not introduce any security issues of the operational level, either.

A guideline to track down and verify a route or an AS path to a valid or a forged root server is described in section 2.

Furthermore, if an ISP or a site operate its own anycast root server, hosts of the ISP or the site using the root server is protected from external forged route.

In addition, if a lot of local root servers share an anycast address, it reduce the effect of distributed denial of service attack on the anycast address.

6. Author's Address

Masataka Ohta
Graduate School of Information Science and Engineering
Tokyo Institute of Technology
2-12-1, O-okayama, Meguro-ku
Tokyo 152-8552, JAPAN

Phone: +81-3-5734-3299
Fax: +81-3-5734-3299
EMail: mohta@necom830.hpcl.titech.ac.jp