Distributing K-Root Service by Anycast Routing of 193.0.14.129

Daniel Karrenberg <daniel.karrenberg@ripe.net>

Document ID: ripe-268
Date: 10 February 2003

Abstract

This memo proposes to distribute the DNS service provided by k.root-servers.net across multiple locations in the Internet topology. It discusses the motivation for and the principles of implementation. A first inventory of detailed issues is provided in an appendix.

Table of Contents

- 1.0 Introduction
- 2.0 Scope of this memo
- 3.0 Proposal
- 4.0 Operational & Funding Models
  - 4.1 Traditional
  - 4.2 Location Fees
  - 4.3 Operated and Funded Decentrally
- 5.0 Implementation Plan
  - 5.1 Initial BGP change
  - 5.2 Move the Cold Standby to Service
  - 5.3 Implement Further Instances
  - 5.4 Spreading Further
- Acknowledgements
- Appendix A - Detailed Issues

1.0 Introduction

DNS root name servers need to be accessible by Internet hosts in order for the DNS to function properly. Accessibility is determined by the ability of the server to handle a given query load and by the connectivity of the server to the rest of the Internet.

The RIPE NCC operates k.root-servers.net (K-root) in the RIPE region in order to help safeguard the quality of the DNS in the Internet and in the RIPE region in particular. The RIPE NCC obtains guidance from RIPE.

For K-root, the connectivity issue has been addressed by placing it at the LINX, a topologically very well connected point. The server load issue has been addressed by deploying successive generations of hardware with increased processing power and by distributing the load locally among a number of machines at different LINX sites. A cold spare system has been available in Amsterdam at all times to provide continuity in case of catastrophic failures at the primary server location. Over the years this set-up has provided very reliable service.

However, issues about differences in connectivity to the service across the RIPE region have been raised repeatedly. A more distributed provision of the service is generally seen as positive because of:

- lower network delays due to shorter paths between clients and servers,
- less dependence on connectivity to a single location,
- better load and DDoS attack resilience because of distributed servers,
- more overall redundancy.

For these reasons numerous organisations have offered to host additional servers operated by the RIPE NCC. So far this has not been considered because the number of unique IP addresses at which the service can be provided is exhausted by currently assigned servers.

2. Scope of this Memo

This memo proposes to deploy multiple servers providing k.root-servers.net name service across the RIPE region, each using the same IP address. This is commonly called 'anycasting'. A detailed description of one implementation can be found in RFC3258. The intention of this memo is to establish the principles of this, inventarise the major issues and request comments from the RIPE community and other interested parties. An initial inventory of detailed issues is provided as an appendix.

3. Proposal

Simply put, the RIPE NCC will provide K-root name service at multiple locations dispersed over the Internet topology but at the same IP address. No DNS client/resolver changes are necessary. The service will appear exactly the same for the users. Normal Internet routing will distribute the traffic among the different instances of K-root.

This will be implemented by installing multiple copies of the current server sets at different points and announcing the current prefix 193.0.14/24 from these points.

The main challenge with this set-up is to ensure consistency:

- **operation**

  All servers will be operated in a consistent way by the RIPE NCC. They will have appropriate out-of-band access path to ensure that the operations centre can access them.

- **monitoring**

  The availability and responses will be monitored by dedicated monitoring systems installed at multiple locations. Also the BGP propagation of the prefix 193.0.14/24 will be monitored constantly by the existing remote route collectors. [http://www.ripe.net/ris/] Users will be able to identify the particular instance they are using by published methods using DNS queries. [draft-ietf-dnsop-serverid]

- **correctness**

  The correctness and authenticity of the root zone data will eventually be guaranteed using DNSSEC. Until this can be deployed the current method will need to be employed: ISPs will need to closely monitor where they route traffic to 193.0.14/24 and from where they accept traffic from that address. The RIPE NCC will publish and maintain a list with the locations of the k-root instances, the BGP autonomous system numbers of their immediate neighbours and any other information that can help ISPs and others to ensure that they reach an authentic instance of K-root. This list will be maintained until appropriate routing security technology is widely deployed.

The RIPE NCC and K-root itself are well suited for this mode of operation. IANA asked the RIPE NCC to operate K-root, the geographic and topological location of K-root was not specified in any way other than "somewhere in the RIPE region". The RIPE NCC was chosen because it is neutral and professional but above all directly accountable to RIPE and its membership.

The location at the LINX was subsequently chosen based on evaluation of the Internet topology in the region and a recommendation by the RIPE DNS working group. Distributing K- root over a number of places is a natural continuation of this policy.

In addition to operating K-root at a remote location the RIPE NCC has considerable experience in operating distributed services. The Test Traffic Measurements Service operates more than 80 machines all over the world to collect performance measurements. Some of these can be used to monitor the DNS service of K-root as well. The Routing Information Service operates 9 remote route collectors at exchange points all over the world to collect BGP routing information. These route collectors will be used to monitor the routing of 193.0.14/24.

At any point in time there is a trade-off between the added benefit of adding more servers and the difficulty of operating them consistently. The optimal number of servers depends on a large number of constantly changing factors. It needs to be evaluated continuously as things progress.

4. Operational & Funding Models

For the purpose of stability and for gaining experience, all instances of K-root will be operated by the RIPE NCC. This ensures smooth transitions, consistency and correctness of root zone data.

After the initial deployment, there are a number of possible operational and funding models.

4.1. Traditional

Traditionally K-root operations have been part of general RIPE NCC activities and thus have been collectively funded by the RIPE NCC membership. This is an appropriate funding model as all members benefit from stable root name service. It is also easy to administer. Difficulties may arise when the number of locations is such that the operational costs increase very significantly. Another important drawback of this model is that the number of additional sites will be limited by available funds and the sites will have to be determined by a selection procedure based on criteria including:

- position in Internet topology,
- position relative to existing K-root instances,
- local operations support,
- operational requirements [rfc2870],
- commitment to fund operations at a later stage.

4.2. Location Fees

The drawbacks of the traditional model can be largely overcome by charging the organisations hosting K-root instances a fee that covers the operational costs. This obviously scales better and requires less of a beauty contest, because the funds available will more closely match the operational costs. It is quite possible that the initial demand will be higher than the operational capabilities of the RIPE NCC.

Also it should be observed that in funding and some other aspects this is exactly the opposite of the traditional model: In the traditional model the RIPE NCC pays facilities management fees to the hosts whereas in this model the hosts pay.

Transition should be planned carefully.

4.3. Operated and Funded Decentrally

In this model anyone who wishes would be able to operate a K-root instance. This model has such serious problems with guaranteeing stability and consistency that it cannot be implemented today or in the near future. The minimum requirement for this operational mode is a zone signing mechanism that ensures consistency and authenticity of root zone data. Implementing this also requires a changed model of service responsibility as it is obviously impossible to hold any one entity responsible for the service. While this may ultimately scale the best, it is extremely premature at this point.

We propose to continue with the traditional model for now and explore other models while gaining operational experience. The associated selection procedures will be executed by the RIPE NCC and guided by the relevant requirements RFCs and the RIPE DNS working group.

5. Implementation Plan

5.1. Initial BGP change

The routing announcements of the current server prefix 193.0.14/24 will change from AS5459 (LINX) to the dedicated new K-root autonomous system number AS25152.

ISPs need to be aware of this and adapt their routing filters accordingly. It is recommended that ISPs who do not already do so take precautions, safeguarding that they receive this prefix from an authentic K-root via a trusted path and that they route traffic to it via trusted paths as far as possible. At the same time an initial set of BGP and DNS service monitors will be deployed.

5.2. Move the Cold Standby to Service

The current cold standby server at the RIPE NCC will be activated as a regular server. AS25152 will be announced also by the RIPE NCC at the AMS-IX. As the server is already available and configured this can be done fairly rapidly. The distribution of the load, BGP routing information and other operational data will be gathered and evaluated. ISPs will have the opportunity to test this set-up and provide feedback. In case of problems the RIPE NCC instance of K-root will be deactivated quickly, returning to the previous service level.

5.3. Implement Further Instances

While monitoring continuously a number of further instances of K-root will be deployed. Currently we expect this to be about 5 additional instances. This number is limited by operational and monitoring capacity. It is important not to deploy more instances than can safely be operated and monitored. Especially the capacity to detect and correct problems in this distributed set-up needs to be carefully monitored.

Those interested in hosting an instance of K.root-servers.net should contact the author.

5.4. Spreading Further

Planning further than this is currently difficult because of lack of experience with distributed K-root operations and possible funding models.

Acknowledgements

The author would like to thank Joao Silva Damas for comments on an earlier draft as well as Andrei Robachevsky and Henk Uijterwaal who provided comments on recent drafts.

Appendix A - Detailed Issues

Using a router or having a server speak BGP?

Withdrawing BGP announcements based on service availability?

Distributed service monitoring: Using current RIPE NCC operated machines all over the world is an easy option. Maybe distributing automatic monitoring software to be run by volunteers is another. What is essentially needed is to try a small number of queries periodically including the query identifying the instance of the server. Then transmitting the results back to a (set of) collection point(s).

Distributed BGP Monitoring: The RIS can be used for this. Coverage should be sufficient.

Information Campaign: What is needed to reach all ISPs that need to know? How long a lead time do they need for the various stages of the deployment?