**Experiment proposal for Macaque Project**

**BII Group, TISF, and WIDE Project**

**1. Introduction**

"One World, One Internet, One Namespace" is the essence for the success of today's Internet. However, as the top level of the unique identify system, the DNS root system, which has been operational for 25+ years, is a pivot to make the current Internet useful. So it is considered somewhat ossified for stability reasons. It is hard to test and implement new ideas evolving to more advanced level to counter the challenges like IPv6-only operation, DNSSEC key rollover, scaling issues, etc. Because any changes to the current production system is not risk-free, a large-scaled Testbed is necessary on which fully test and study on the impact can be done. It is possible to pre-evaluate such a change in a laboratory test, however, the coverage is very limited. In order to make the test more practical, involvement of users’ environment which varies highly diverse manner is also necessary to study the effect of the change in question.

(following highly related to the discussion in WIDE Camp should be rewritten or move to the Appendix)

Trying to answer these questions, an unconference on the topic of DNS Root testbed was held in the afternoon of March 14th in WIDE CAMP. A brainstorming was invoked between participants including the current C root and M root operators. The general consensus was formed and listed as follows:

1) The participants are in favor of setting up a new group of root operators who are dedicated to build a paralleled research and experimental live root system compared to the existing production one.

2) WIDE, FSI and BII are initiators and funders of Macaque project in which they have responsibility to propose, examine and implement experiments in this project. They also play a key role in selecting participants, send invitations and dealing with public relationship.

3) Possible research agenda covering several aspects but not limited to: IPv6-only operation, DNSSEC experiment, Renumbering issue, the scalability issues, multiple zone file signers.

4) Participants will be invited both on server side or client side. Measurement study will be done on both side to issue out useful analysis as output of our testbed

5) The experiment will be conducted until 2018-12-31 for the expected experiments.

6) We named the project as Macaque Project: A Testbed for Advance Root Services. The word “Macaque” has symbolical meaning to our activities.

7) As to teamwork tools, google group and google doc are utilized to make our collaboration more close and efficient. In addition periodic meeting and gathering is necessary for project management and information sharing.

Note that the root testbed is supposed to be built following the structure of the current 13 root server with the feed of zone file from IANA. To be clear that the research root is strictly in line with the IANA namespace except for the APEX records which is necessary to introduce a new group of root operators for the experiment. The DNSSEC signature associated with each resource record set are altered accordingly. This means participants are required to install the testbed specific DNSSEC key.

To make this project more public mailing list and some kind of wiki is necessary for interested parties and individuals.

**2. Overall testbed structure and “Who should attend”**

The testbed consists of server side and client side, as seen in the current DNS hierarchy.

The server side consists testbed-Root DNS servers which will be operated by BII, TISF, and WIDE Project (this list is just tentative; added when necessary to provide better coverage to the participants scattered globally).

The client side consists of DNS resolvers (full or partial) and end resolvers (end hosts). Client side environment will be provided by each participant. Each system doesn’t have to be an entire company but a small division of a company still fit better.

Who should attend:

* DNS software implementers (full resolvers, partial resolvers, authoritative servers, or even stub resolvers)
* Developers of CPE devices, IoT devices, …, which may include DNS functions to test their implementations with various different shape of the Root DNS servers.
* Middlebox developers (security devices, NAT devices, etc) even no direct interaction with DNS is implemented as these boxes could interfere the smooth flow of DNS packets especially the size of the packet is large, even fragmented in IPv4 and/or IPv6.

who should attends and what for

traffic monitor / records

clock synchronization requirements

**3. Problem Statement (or research description)**

This section briefly introduce what concrete technical problems we are going to focus on. It serve as a problem statement. Aiming to using our testbed to study the problem by experiment, each experiment will be followed with a dedicated document fully describing how to conduct the experiment and what the expected outcome is.

**3.1** **IPv6-only Operation**

In the field of IPv6 development, dual stack is view as a practical approach to quickly deploy IPv6 services coexistent with IPv4 infrastructure. It achieved success in many ISPs and ICPs. Although IPv6 network and IPv6 enabled application deployed, many of them still rely heavily on IPv4 infrastructure. Everyone expects IPv4 will fade gradually, but no one knows whether Internet can survive without IPv4 or any problems exists for end-to-end IPv6-only communication. In addition, operation on the dual stack network involve additional/high complexity and cost for operators. It is surely desirable to run a single stack network and system for the future network.

DNS is the vital part of Internet infrastructure whose IPv6-only capability is a key to support end-to-end IPv6-only communication. Although the inherent feature of DNS protocol give it more independence on IP protocol, DNS operation in IPv4 and IPv6 is not actually the same. IPv6 with larger MTU which may relax the packets sizes issue in IPv4 context. Packet fragmentation in IPv6 may cause problems following current practice. Middle box and network environment may be different. All in all, IPv6 only operation experience is very important for DNS operation especially on Root level.

**3.2** **DNSSEC Key rollover and even algorithm rollover**

DNSSEC, as a collection of new resource records and protocol modifications that add data origin authentication and data integrity to the DNS[RFC4035], has become a dispensable part of Domain Name System. Key roller, as a procedure to update the public key in resolvers, has been a significant issue in DNSSEC. Nowadays, IANA rolls ZSK every six week but for reliability issue, KSK is never rolled. Thus, the way of rolling KSK and the effect of rolling key (including ZSK and KSK) frequently are never fully researched. It is worthwhile to to test KSK rollover using RFC5011 to synchronize the validators in a live DNS root system

In addition, in current system DNSSEC key is 1024 bits using RSA. The effect of using key with more bits is never tested because longer key will enlarge DNS answer package with DNSSEC, which may cause a lot of operational problems. It is valuable to observe the effect of changing keys’ bits in test environment. Encryption algorithm such as ECC, as another factor that would also affect package size, should also be tested.

**3.3 Renumbering issues**

Nowadays Internet users or enterprises may change their network providers. As a result their Internet numbers for particular servers or services like IP address and AS numbers may change accordingly which is called renumbering their networks and servers. It is likely that root operators may change their IP address for root servers as well. Since there is not dynamic update mechanism to info that changes to resolvers and other internet infrastructure relying on root service, the renumbering issue of root server become a fragile part of the whole system.

Fortunately the renumbering of root server is rare and the system is relatively stable for many years, except for a handful of cases that root servers were renumbered. It is observed that while new root address is adopted for a long time, old address still got the traffic. It is interesting to test how long the old address still gets traffic and what kinds these traffic are.

Due to the worry about DNS hijacking, there is never a possible way to dynamic me update new root servers’ address to hint file. However, with the deployment of DNSSEC, it is possible to prevent resolvers from DNS hijacking attack. Thus, designing a mechanism to dynamic update root servers’ address to hint file like key rolling in DNSSEC becomes possible. Hopefully via this testbed, we can come up with a mechanism to specify the secure hint file rollover in priming process.

**3.4** **Scalability problem of Root server system**

In the current production network, there are 13 root servers (root letters) with more than 400 root instant serving the global internet users. One observation is that the distribution of root server system is not geographical-evenly distributed. It may cause fatal internet outage due to the failure accessing the root services. It is particular the case for some networks and regions which is not well connected with only a fiber to outside for example.

To distribute the root server more pervasively, there are several approaches. Most discussions now are focused on how to increase the number of machines and capacity of this system other than increasing the number of operators (increase letters). Because the idea of increasing root operators not only challenge the current DNS packet size limitation of 512B, but also (more importantly) may introduce controversial topics such as how to select new root operators, how many operators is good enough, is it better to reduce the number 13? So there is an interesting research question here: how many root name servers is too many, and what are the limitations, and what is the optimal number of root name servers?

For a research point of view, given the root system containing as many letters as possible, there are some tradeoff to be considered carefully. 1) more root servers surely increase resilience for particular users, but 2) the resolver will perform round trip measurement and keep state for each potential name server which may introduce more cost than usual; 3) also with that many servers, error theory predicts that some number of them will be unreachable, causing retries that add to the total number of referral events required to locate the closest (by SRTT) root server.

**3.5** **Multiple zone file signers**

Now the IANA system only has one signer (Verisign) which introduce single point of failure, we do not know multiple signer will work. Intuitively we intent to introduce more players sharing the same key to maintain and sign the zone.

One possible way is that WIDE, FSI and BII can in turn generate the key, share the key with others, sign the zone independently and distributed the zone file to other root server. Root server will receive three copy of signed zone files which can be validate with the same public key. Root operator can decide which copy is used and others two are for backup.

**4. Who and How to participate Macaque Project**

Section 2 depicts the project with players who will compose the testbed, participate and benefit from the experiments. Since each player in this testbed is important to the success of Macaque Project, it is necessary to formally define the specific roles, functions, interfaces and data flows for each player. This section firstly introduces the basic outline of Macaque Testbed, and then describes the roles in the testbed, how they can participate, cooperate and fulfil their function. Note that one participator is not limited to play one role in the project, he can play several roles as well.

**4.1 Outline of Macaque Testbed**

There are some basic technical configuration of this testbed which should be familiar by any participants in Macaque Project.

1. A set of Zone files of Macaque Testbed. It contains three parts:
   1. a new hint file corresponding to new root servers
   2. a new zone file which is same with IANA zone file expert for APEX records
   3. a new Zone DNSSEC Trust Anchors which maintains the DS and KSK;
2. Zone synchronization. An routine to fetch the latest zone file periodically from IANA
3. IPv6-only environment. All components of the root system including DM, root server, monitoring system are supposed to run in IPv6-only network. Resolver is not required to run in IPv6-only environment unless it act as a forwarder turn to upstream dual-stack recursive server.
4. Use normal fully qualified domain names, not single-letter names having a shared parent domain name.
5. Share one unique key for the new group of root servers and our data will be DNSSEC signed
6. Multiple DMs for root file distribution (act as VeriSign does) which has different IP address from Root servers.
7. Rapid Key rollover time (ZSK twice a month /KSK four times a year)
8. tsig keys for zone transfer
9. Monitor and collect all the traffic data from both client and server sides. The testbed provide two interface to access the data. One is the realtime DNS data information. The other is the historical DNS data information.

to make it more clear there is a table to compare our Macaque Root Testbed with the existing system in production network which shows the clear difference and resemblance:

Table 1 The Comparison of Two Root system

|  |  |  |
| --- | --- | --- |
|  | Existing Root system | Macaque Root Testbed |
| DNSSEC key | Deploy DNSSEC in root zone with IANA’s KSK and Verisign’s ZSK | Deploy DNSSEC in root zone with a new set of ZSK/KSK. |
| Key rollover | Rollover ZSK manually every 6 weeks. Do not rollover KSK . | Rollover ZSK/KSK automatically following RFC5011 with more rapid frequency (2 weeks for ZSK, 3 months for KSK) |
| IPv6 | Dual-stack for A,B,C,D,F, H,I,J,K L,M. Only E and G are IPv4 only. | IPv6-only |
| number of root servers | 13 | No limitation in theory. But for performance consideration as a live root system the number will be optimized. |
| root server operator | the current 12 operation :<http://root-servers.org/> | The selection of root server operators in Macaque project will be much based on geographical consideration. |
| zone file | updated by ICANN/IANA, verificated by NTIA, and signed by Verisign | strictly in line with the IANA namespace except for the APEX records which is necessary to introduce a new group of root operators for the experiment. |
| hint file | code the 13 root servers inside | code the root servers of Macaque project in side |
| scaling issue | Deploy root instance via BGP peering | increase the number of root server as well as root instance. |
| DNS protocol | do not change | May introduce some new transmission protocol and extension for experiment. |

**4.2 Role Definition in Macaque Project**

**IANA function operator**

Generally speaking root service is the DNS lookup service for TLD (gTLD and ccTLD). The Root zone contain the latest delegation information for each TLD. The TLD operators submit change requests occasionally to ICANN as the current IANA function operator. Accordingly ICANN will checks that the change requests meets policy and technical requirement and confirm consent from the appropriate parties. if issues are found, ICANN clarifies with the TLD operator. Then ICANN forwards the request to NTIA for verification and to Verisign who is the current root zone maintainer. That’s all the processus done by IANA function operator.

In Macaque Project, ICANN still plays IANA functions whose effort is a key to keep “One namespace” and update it. The different is that Root Zone Maintainer (RZM) of Macaque Project will actively fetch the latest zone file from ICANN before it is signed. (or change the signed file into original one). Note that no one will change the namespace maintained by IANA except for the APEX records which is necessary to introduce a new group of root operators for the experiment. in addition if certain experiment require changes to the TTL of root zone, the changes will be reflected to the zone file accordingly.

**Root Zone Maintainer (RZM) and Distribution Master (DM)**

In production network the role of RZM and DM is played by Verisign who takes the responsibility of managing the Zone Signing Key (ZSK), generating (or updating) root zone file with its digital signature and sending the signed zone file to other Root Server Operator (RSO). In the current situation, Verisign is the single party who perform all the procedures as RZM and DM which introduce latent single point of failure.

In Macaque Project, the role of RZM and DM will play by WIDE, FSI and BII the founders of this projects. They will actively and periodically fetch the zone file from ICANN/IANA to keep one unique namespace for TLDs. In the meanwhile, they will keep one unique set of ZSK/KSK and rollover the key more rapidly than the current practice. Idealy each one of them can take charges for a certain period of time, maintaining hint file, downloading the zone file, generating the keys, sign the zone file and share the zone data with other two. In the beginning BII will firstly take charge the role of RZM and DM to quickly setup and validate the system according to the basic of requirement of the testbed.

**Root Server Operator (RSO)**

In production network, 12 operators run 13 root servers (from A to M). They are Authority server for hosting the zone file. Their domain and IP addresses (both IPv4 and IPv6) are listed in the hint file which is publicly released by ICANN/IANA and configured in hard code by default in some DNS software implementation like BIND and PowerDNS. When a resolver bootstraps it start priming query to the addresses listed in hint file or hard code. So the hint file and configuration becomes a switch to choose which root system it is mounted on.

In Macaque Project, the role of RSO is very important who will be solicited to join the project. To start with, only WIDE, FSI and BII take the role of RSO. A new hint file will be maintained including the information of all RSOs selected. Each RSO will provide a dedicated domain (FQDN) and IPv6 address which will be included in the hint file. Each RSO will become the member of **Macaque Project Committee(MPC)** which commit to build the project and contribute on Hosting the root, monitoring the traffic and developing its own Root server network. If a RSO want to change its domain or address or quit the project, it will trigger the renumbering process to handle this event. Note that renumbering process can be simply changed by RZM or complex with a predefined procedure

There are some document regarding how to build the root server[F-root][RFCs]. and also there are technical and performance requirements of RSO in current production network is documented by RSSAC [rssac 001,002]. Note this documents only serve as reference, but variety of this system is the key to robustness of the system. So the root server operator can decide independently how to build the root system in Macaque projects.

**DNS Resolver (RDNS) : Stub or Recursive Server (IPv6-only or dual stack?)**

The function of RDNS in Macaque project is of no difference from the one in production network except that the invited RDNS as a client will configured with the new hint file of Macaque project. The new hint file will direct the RDNS to mount itself to the Macaque root testbed.

It’s worthwhile to mention that RDNS is the important interface to build the real users environment for Macaque Root testbed. The players attending the project will take the role of RDNS in the first place :

* For DNS software implementer it is straightforward to deploy full resolvers, partial resolvers and stub resolvers with its own implementation. Then only need to configure these resolvers with the new hint file maintained by Macaque Project. If some public open recursive server want to join our project, it will be listed as open DNS server of Macaque which will help to introduce more traffic in to this testbed.
* For developer of CPE devices , IoT devices which may include DNS functions, they can test their implementation
* For Middlebox developers (security devices, NAT devices, etc) , they can connect Macaque testbed with their testbed by setting up a RDNS mounted in Macaque root. Although even no direct interaction with DNS is implemented, these boxes could interfere the smooth flow of DNS packets especially the size of the packet is large, even fragmented in IPv4 and/or IPv6.
* For normal users who are interesting in attending the project, like a lab of a campus or a branch of institute or individuals, they can just configure their own local recursive server or make the public recursive server of Macaque Project as their recursive server.

**Traffic and Data Collector (TDC)**

Running code and traffic data are key driving force for Internet. They are also the key to our project, especially the traffic data which are valuable during the experiments done in the testbed. Like the DITL activities in production network, it is important for Macaque Project to conduct a large-scale, simultaneous traffic data collection events with the goal of capturing datasets of strategic interest to member of the project and other researchers.

To achieve this It is suggested to run passive DNS sensor and upload data to SIE (Security Information Exchange). FSI will provide remote access free of charge for simultaneous traffic data. BII will run archive and provide remote access free of charge.

**Traffic and Data Analyzer (TDA)**

TDA is the role who consumes the data from TDC. In principle TDA is free to access the traffic data and log information, but if there were an incentive policy which encourages potential TDA to contribute the project by taking the role of RDNS in the same while, it would be much better to promote the macaque project. Atlas in RIPE NCC is an typical example.

**Experiment Proposer (EP)**

EP play an active role to propose specific experiment proposal. Each proposal will be firstly vetted by MPC (Macaque Project Committee) and then implement in the testbed.

<more suggestion needed here to specify the procedure to conduct an experiment >

**Macaque Project Committee (MPC)**

Composed by RSO, RZM and DM, MPC serves as a decision-making role in Macaque Project such as making plan, examining experiment proposals etc. As the founder of Macaque project WIDE, FSI and BII are the co-chairs of that committee.

**Macaque Project Technical Caucus (MPTC)**

MPTC is a group of invited expert and researcher who are instereted and help to propose and conduct experiment. To maintain a good public relationship, it’s worthwhile convening a group of experts who advocate our activities.

**4.3 How to coordinate the project**

Now there are many tools to coordinate the work and people even in different time zone. We prefer using GitHub for public release, such as Macaque Project’s wiki, documents, zone file and public key. Google Doc is perfect for internal joint paper work. In addition Google group will be utilized as a mailing list of the project. Particularly we commit to running three mailing lists:

1) Authority operators(root and DM),

2) Recursive operator (RDNS+#1),

3) interested parties (+#2,#1).

Before Macaque Project public released, founders of this project should commit to meet every week to finish the experiment proposal, verify the root testbed internally, brief to some interested participants and prepare for public release. For the daily operation we should commit to a monthly newsletter sent to list #3 above, showing current statistics, recent changes, and upcoming changes. We should commit to a quarterly "jitsi meeting", inviting everyone on #3 to join us.

**4. The Development Plan**

This section introduces the plan and schedule how to build the project in several milestones. It’s aiming to provide founders of Macaque Project clear path how to start the project by each step and the goal in each step. Normally a development plan of a project include several basic part: participants, action plan, resource needed, risk analysis and expected result (or benefit). The previous sections cover most part of participant and role. in this section we will focus on action plan. The action plan will composed as a ToDo list which are describe as follows:

1. Create, review and approve document:
   1. To fully discuss and complete the Macaque Project proposal which is the first document to focus on project description, roles and motivations. The first draft of this document should be issue out before 10th April;
   2. To document specific experiment proposal for the research topic which will focus on design, data flow, monitoring and expected output; it should be proposed, vetted and public released. The proposed deadline for first experiment proposal will be the end of April.
   3. Prepare Instructions for participate both for server side and recursive side. The proposed deadline for first experiment proposal will be the end of April.
2. At the same time with step #1, to construct initial test bed.
   1. BII will setup the first distribution master and share the zone file and key with other distribution master; It is suppose to finish before 10th april;
   2. Then, BII will set up first authority server and recursive server to test it works in IPv6 pure environment with rapid key rollover settings; It is suppose to finish before 17th april;
   3. Other authority and recursive server will be set up after they got the zone file and key.It is suppose to finish before 17th april;
   4. To setup the monitoring system, preferable using SIE from FSI and some kind of log collector; It is suppose to finish before end of April;
3. After finishing #1&#2, prepare for announcement in 18th April (A conference host by BII in Beijing, ICANN CTO David will be there)
   1. To create and review the announcement text with the internal approvals at WIDE, FSI, and BII; First draft before 24th April, Deadline 8th May;
   2. To create, review and approve authority server invitation, both the candidate and invitation letter; First draft before 24th April, Deadline 8th May;
   3. Leave at least one week for WIDE board member to prove this project.Necessary document should be prepared before the end of April;
   4. WIDE, FSI and BII will set up website individually to introduce Macaque project. They can link each other and point to the same GitHub page. It is suppose to finish before 10th May;
   5. To send invitation to the candidate just one day before announcement day, 17th April;
   6. On 18th April, send announcement to dnsop and dns-operation;

1. Daily operation (after the announcement) TBD
   1. Receive offers of root server operation -- review and approve
   2. Receive inquiries about project -- create "canned" answers, review and approve, send responses
   3. Operate and monitor the system on a daily basis
   4. At appropriate intervals, renumber servers or roll keys or whatever
   5. Receive, review, approve proposals for experiments
   6. Coordinate experiments with their proposers
   7. Issue monthly status reports
   8. Hold "jitsi" meetings every six weeks
   9. Hold in-person meetings twice per year
   10. Publish final report