f.root-servers.net

Atelier ccTLD ISOC
Dakar, Décembre 2005

Présenté par Joe Abley
traduit par Alain Patrick AINA

Les fondamentaux

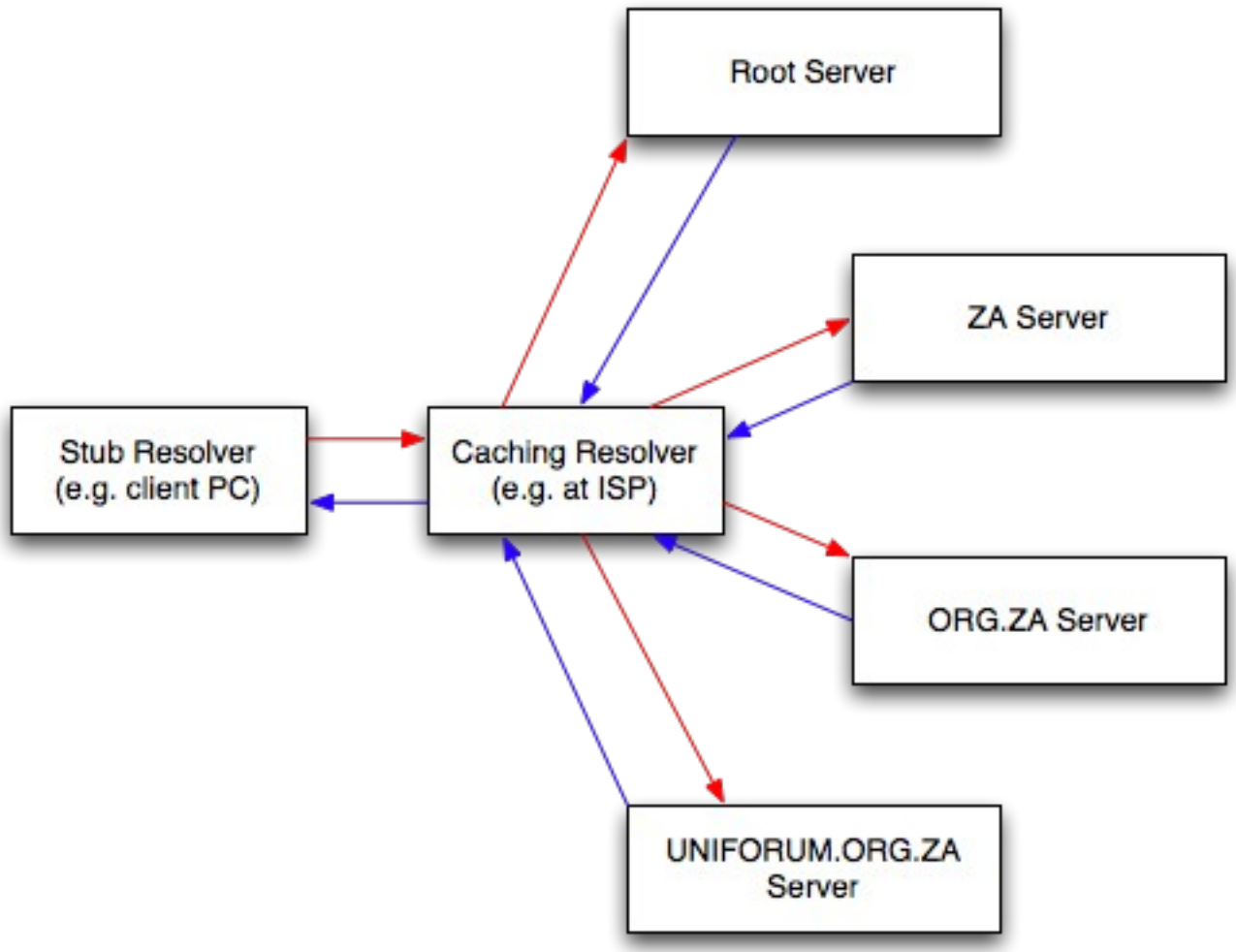
DNS

- Le système de noms de domaine est une grande base de données d'enregistrements de ressource
  - Globalement distribuée, cohérente, évolutive, fiable, dynamique
  - Fait correspondre les noms à divers autres objets
- Le DNS permet l'utilisation des noms pour trouver des ressources sur l'Internet, au lieu des nombres

Composantes du DNS

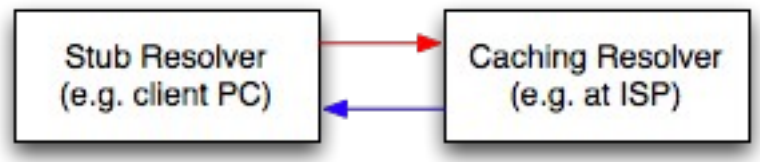
- Un espace de nommage
  - hiérarchisé, comme la structure d'un arbre
  - Des noms séparés par des points
- Les serveurs de nom
  - Les serveurs qui répond aux requêtes des clients, et rendent les données disponibles
- Resolvers
  - Les clients qui envoient des requêtes

www.uniforum.org.za



www.uniforum.org.za

- Les réponses déjà dans le cache sont envoyées directement, sans une recherche récursive
- Les données expirent du cache quand elles vieillissent



Serveurs racine

- Chaque serveur récursif a besoin de savoir comment atteindre un serveur racine
- Les serveurs racine sont des points d'entrée bien connus pour le DNS
- Il y a 13 adresses de serveurs racine, situés à différents endroits, gérés par différentes personnes
- La zone racine est publiée par l'IANA

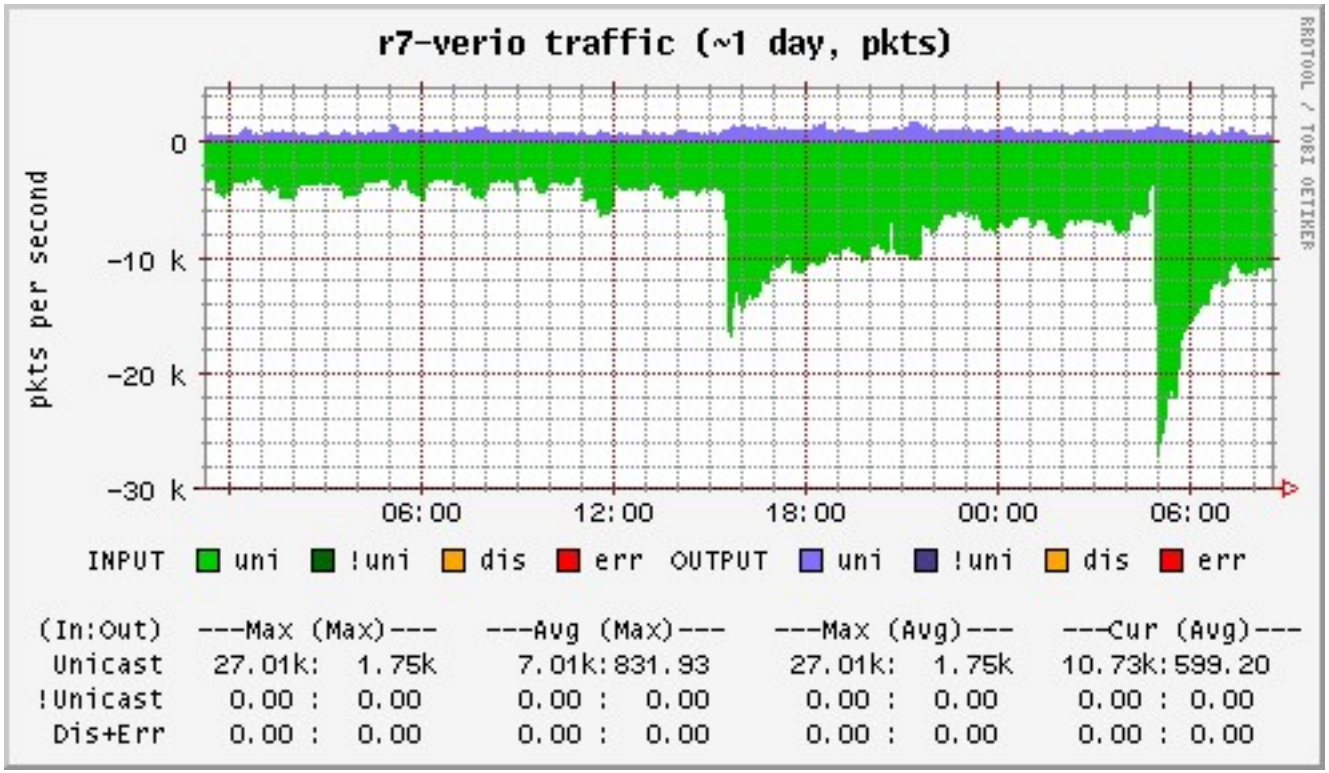
Les serveurs racine

A.ROOT-SERVERS.NET	Verisign Global Registry Services	Herndon, VA, US
B.ROOT-SERVERS.NET	Information Sciences Institute	Marina del Rey, CA, US
C.ROOT-SERVERS.NET	Cogent Communications	Herndon, VA, US
D.ROOT-SERVERS.NET	University of Maryland	College Park, MD, US
E.ROOT-SERVERS.NET	NASA Ames Research Centre	Mountain View, CA, US
F.ROOT-SERVERS.NET	Internet Software Consortium	Various Places
G.ROOT-SERVERS.NET	US Department of Defence	Vienna, VA, US
H.ROOT-SERVERS.NET	US Army Research Lab	Aberdeen, MD, US
I.ROOT-SERVERS.NET	Autonomica	Stockholm, SE
J.ROOT-SERVERS.NET	Verisign Global Registry Services	Herndon, VA, US
K.ROOT-SERVERS.NET	RIPE	London, UK
L.ROOT-SERVERS.NET	IANA	Los Angeles, CA, US
M.ROOT-SERVERS.NET	WIDE Project	Tokyo, JP

Modes de dysfonctionnement du DNS

Les défis sur la racine

- Il y a eu un certain nombre d'attaques contre les serveurs racine
- Des DDoS peuvent générer beaucoup de trafic, et rendent les serveurs racine inaccessibles à plusieurs personnes
- Un temps d'inaccessibilité prolongé pourrait conduire à un dysfonctionnement à une grande échelle du DNS



C’est une jungle là-bas

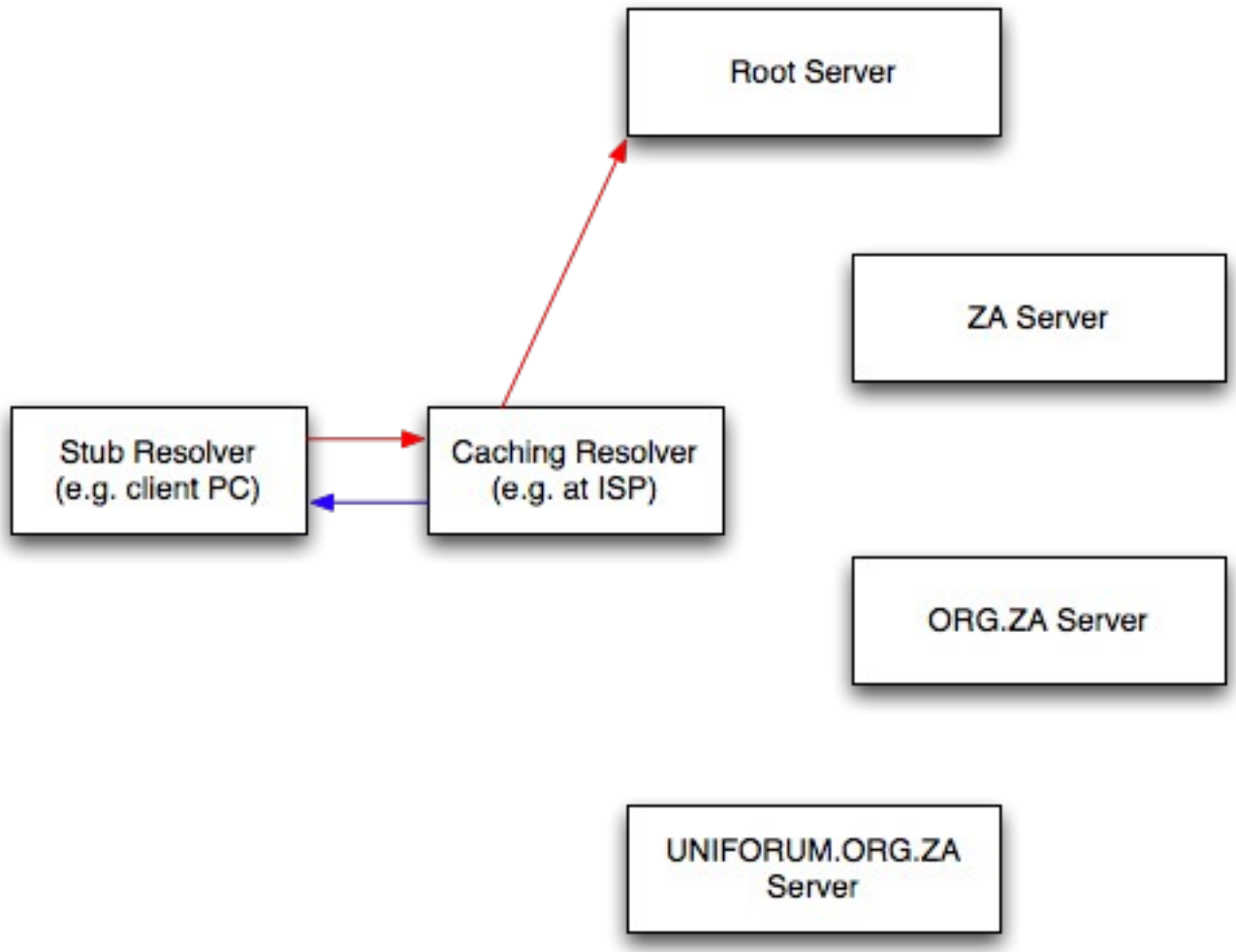
Dysfonctionnement global du DNS

- La Probabilité de dysfonctionnement du DNS entier est faible
  - Les plus importantes données du DNS (enregistrements fréquemment demandés) sont mis en cache, généralement avec de grands TTL
  - Les serveurs racine sont gérés par des entités indépendantes et sous une sécurité substantielle
  - Les attaques coordonnées contre les serveurs racine tendent à être examinées rigoureusement

Dysfonctionnement régional du DNS

- Si une région devient déconnectée de l'Internet, ou souffre d'un manque d'accès prolongé aux serveurs racine pour d'autres raisons, le DNS pourrait ne pas fonctionner dans cette région
- Les problèmes affectant de petites régions n'attirent pas la même attention comme les problèmes affectant le réseau global
- Les dysfonctionnements régionaux du DNS sont plus probables qu'un dysfonctionnement global

www.uniforum.org.za



Perte de réseau

- Plusieurs pays dépendent d'un ensemble relativement non diversifié de réseaux extérieurs pour atteindre le reste du monde
  - Un câble sous-marin; un opérateur satellite
  - Un point commun de terminaison dans un hôtel des PTT quelque part
  - Un réseau international saturé, et qui devient inutilisable quand inondé par du trafic non désiré

Le serveur racine F distribué

f.root-servers.net

- A une seule adresse IPv4 (192.5.5.241)
- A une seule adresse IPv6 (2001:500::1035)
- Les requêtes envoyées à ces adresses sont routées vers différents serveurs, en fonction d'où vient la requête
- Ce comportement est transparent aux équipements qui envoient les requêtes à F

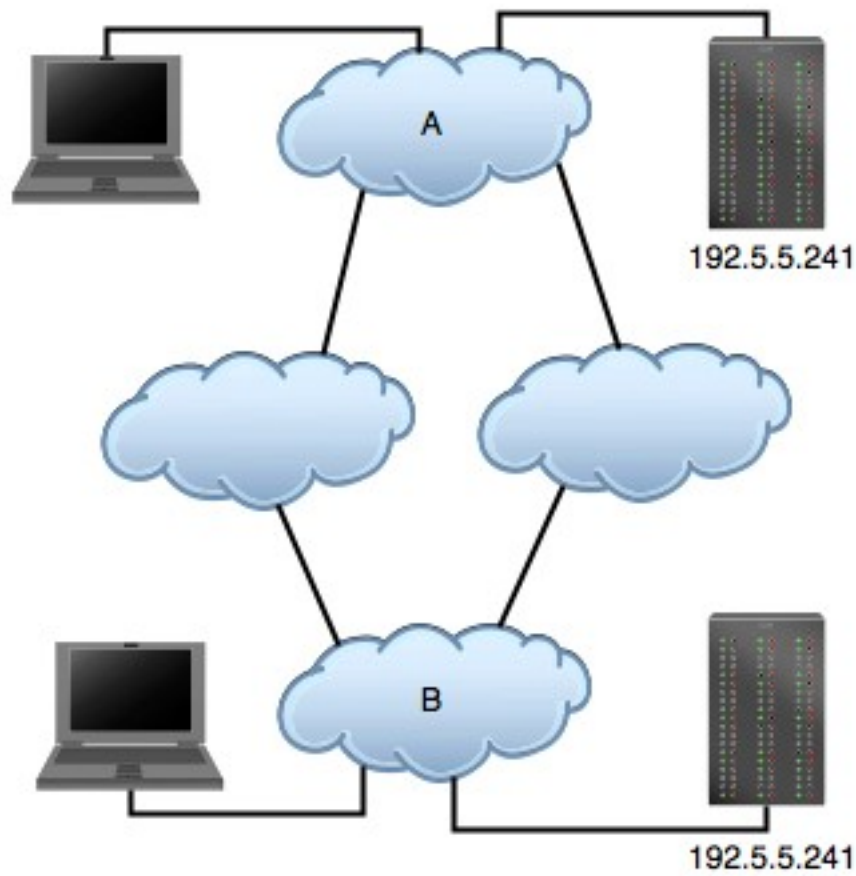
Unicast, Multicast

- La plupart du trafic sur l'Internet est unicast
  - Les paquets ont une seule destination
- Certains trafics sont multicast
  - Les paquets sont destinés à plusieurs destinations

Anycast

- Le trafic vers f.root-servers.net est anycast
  - Les paquets sont destinés à une seule instance de F, mais différentes requêtes (de différentes places) peuvent atterrir sur différentes instances
  - Anycast est identique à unicast de la prespective du client envoyant une requête

Routage Anycast



Hiérarchie Anycast

- Certaines des instances du serveur racine F fournissent des services à l'Internet entier (Instances globales)
  - Des instances très grandes, bien connectées, sécurisées et hyper administrées
- Les autres servent une région particulière (Instances locales)
  - Plus petites

Hiérarchie Anycast

- Le routage de chaque instance locale est organisé de telle manière qu'elle ne servent pas des clients d'autres régions du monde dans les conditions normales
- Pour plus de détails, voir:
  - **http://www.isc.org/tn/isc-tn-2003-1.html**

Modes de dysfonctionnement

- Si une instance locale échoue, les requêtes vers F sont automatiquement routées vers une instance globale
- Si une instance globale échoue, les requêtes sont automatiquement routées vers une autre instance globale
- Les dysfonctionnements catastrophiques de toutes les instances globales conduisent à une continuité de services fournis par les instances locales au sein de leur régions

Modes de dysfonctionnement

- Si une région perd la connexion internationale (e.g.coupure du câble sous-marin), les accès aux serveurs racine sont préservés par l'instance locale
- Une fois, la racine accessible, les autres serveurs locaux sont également accessibles reachable (e.g. Serveurs ZA , serveurs ORG.ZA)
- Une fois les serveurs des noms de domaine de premier niveau (TLD) sont accessibles, le trafic interne au pays peut aboutir

Modes de dysfonctionnement

- Une attque de déni de service contre F lancée en dehors de la région est invisible par les utilisateurs de cette région
- Une attque de déni de service lancée contre F au sein de la région est invisible par les autres régions
- Une attaque de déni de service distribuée à grande échelle causera des dégâts proportionnels à la taille de la région (probablement, peut-être)

Triangulation

- Beaucoup d'attaques de déni de service utilisent du trafic à adresse source falsifiée
  - Très long à retracer à travers un réseau
  - Les attaques s'arrêtent généralement avant que les investigations ne se termine
- Observer les réactions relatives d'une instance locale à une attaques peut aider à en identifier la vraie source

Logistique et administratif

Sponsor

- ISC est une organisation à but non lucratif
- Equipments, colo, connectivité réseau pour les instances distantes sont payés par des sponsors
- Tous les équipements sont exclusivement gérés par les ingénieurs de ISC
- Le sponsor couvre les frais de fonctionnement de l'instance distante de ISC

Etat du déploiement

Instances globales

- Palo Alto
- San Francisco

Instances locales

- Amsterdam, Barcelona, Lisbon, Madrid, Moscow, Munich, Paris, Prague, Rome
- São Paulo
- Los Angeles, Monterrey, New York, Ottawa, San Jose, Toronto
- Beijing, Dubai, Hong Kong, Jakarta, Osaka, Seoul, Singapore, Taipei, Tel Aviv
- Auckland, Brisbane
- Johannesburg

Instances locales

- Amsterdam, Barcelona, Lisbon, Madrid, Moscow, Munich, Paris, Prague, Rome
- São Paulo
- Los Angeles, Monterrey, New York, Ottawa, San Jose, Toronto
- Beijing, Dubai, Hong Kong, Jakarta, Osaka, Seoul, Singapore, Taipei, Tel Aviv
- Auckland, Brisbane
- Johannesburg, **Nairobi**

Le F de Nairobi

Importantes statistiques

- Physicquement situé au même endroit que le switch de KIXP
- Une connexion à 100 Mbit/s au KIXP
- Deux connexions redondantes et de faible capacité via deux FAI indépendants pour la gestion, les statistiques, et les transferts de zone
- Un Cluster de deux serveurs de nom partageant la charge

Utilisation du F local

- Depuis nairobi:
  - **`traceroute f.root-servers.net`**
  - **`dig @f.root-servers.net hostname.bind chaos txt`**

Avant...

```
[halibut:~]$ traceroute f.root-servers.net
traceroute to f.root-servers.net (192.5.5.241), 64 hops max, 40 byte packets
 1  router.cctld.or.ke (196.216.0.62)  1.945 ms  7.147 ms  1.165 ms
 2  196.216.66.5 (196.216.66.5)  44.967 ms  23.918 ms  12.420 ms
 3  217.21.112.4.swiftkenya.com (217.21.112.4)  5.141 ms  9.491 ms  5.791 ms
 4  193.220.225.5 (193.220.225.5)  8.919 ms  5.708 ms  5.898 ms
 5  no-nit-tn-7.taide.net (193.219.192.7)  538.820 ms  539.738 ms  550.056 ms
 6  no-nit-tn-5.taide.net (193.219.193.145)  540.073 ms  551.002 ms  536.818 ms
 7  pos5-1.gw3.osl2.alter.net (146.188.39.1)  535.738 ms  536.197 ms  534.790 ms
 8  so-3-0-0.xr2.osl2.alter.net (146.188.15.97)  535.701 ms  542.140 ms  543.969 ms
 9  so-4-2-0.tr1.stk2.alter.net (146.188.15.61)  541.221 ms  545.562 ms  544.435 ms
10  so-7-0-0.ir2.dca4.alter.net (146.188.11.226)  653.929 ms  652.082 ms  649.199 ms
11  so-1-0-0.il2.dca6.alter.net (146.188.13.45)  658.517 ms  652.177 ms  664.978 ms
12  0.so-0-2-0.tl2.sac1.alter.net (152.63.0.190)  887.784 ms  739.093 ms  717.126 ms
13  0.so-1-3-0.xl2.pao1.alter.net (152.63.48.181)  718.044 ms  720.835 ms  727.418 ms
14  pos1-0.xr2.pao1.alter.net (152.63.54.78)  717.283 ms  716.201 ms  714.212 ms
15  188.atm7-0.gw10.pao1.alter.net (152.63.53.21)  778.208 ms  731.906 ms  832.482 ms
16  isc-pao-gw.customer.alter.net (157.130.205.230)  717.801 ms  712.912 ms  712.718 ms
17  f.root-servers.net (192.5.5.241)  743.804 ms  721.633 ms  746.818 ms
[halibut:~]$
```

… et après

```
[halibut:~]$ traceroute f.root-servers.net
traceroute to f.root-servers.net (199.6.6.14), 64 hops max, 40 byte packets
 1  router.cctld.or.ke (196.216.0.62)  244.241 ms  1.159 ms  1.099 ms
 2  196.216.66.5 (196.216.66.5)  8.678 ms  4.942 ms  31.862 ms
 3  80.240.202.54.swiftkenya.com (80.240.202.54)  22.455 ms  15.803 ms  14.864 ms
 4  198.32.143.125 (198.32.143.125)  40.770 ms  7.192 ms  7.786 ms
 5  f.root-servers.net (192.5.5.241)  10.906 ms  10.894 ms *
[halibut:~]$
```

Sponors

- KENIC