f.root-servers.net

ISOC ccTLD Workshop
Nairobi, Kenya, 2005

The Basics

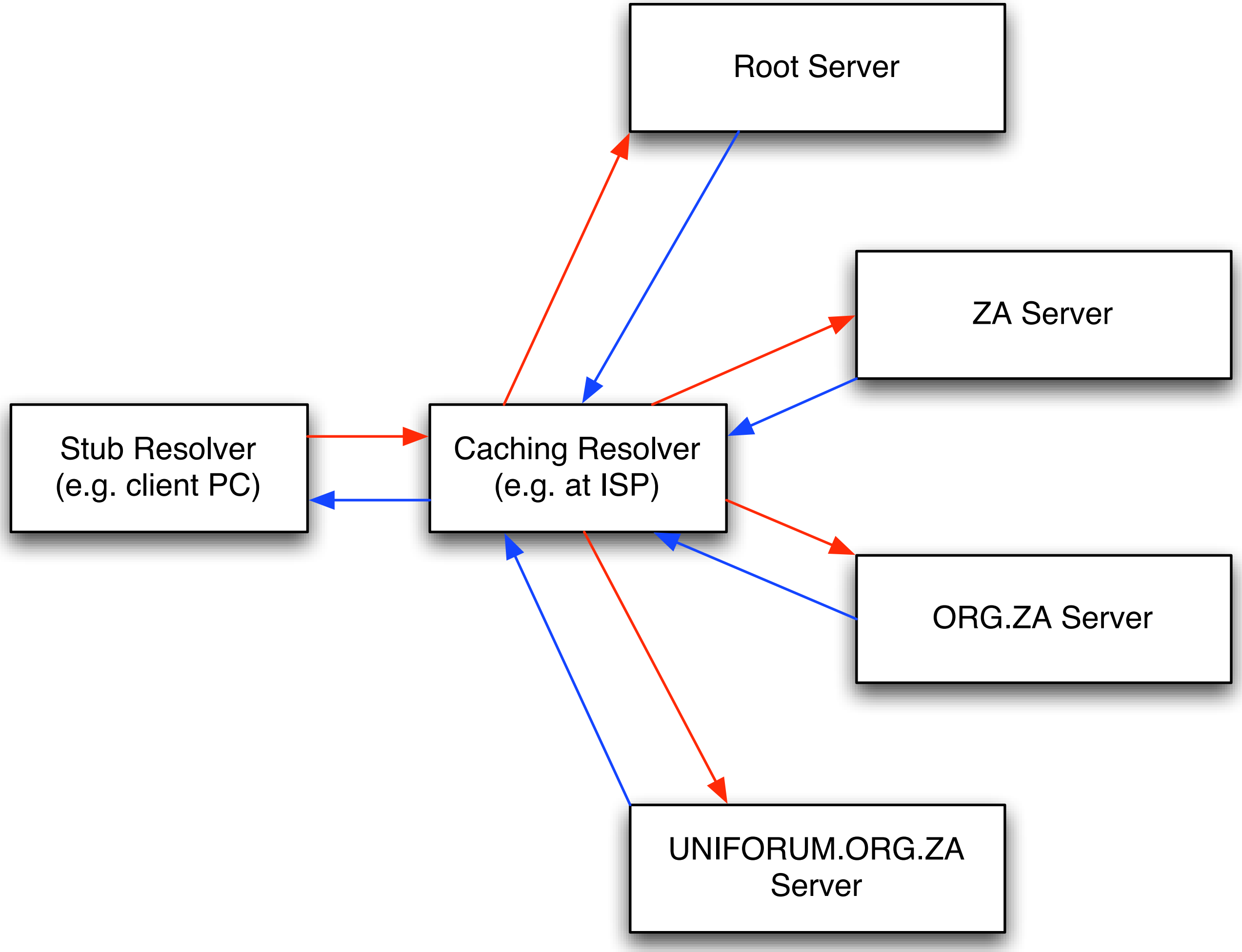
DNS

- The Domain Name System is a huge database of resource records
  - globally distributed, loosely coherent, scaleable, reliable, dynamic
  - maps names to various other objects
- The DNS allows people to use names to locate resources on the Internet, instead of numbers

Components of the DNS

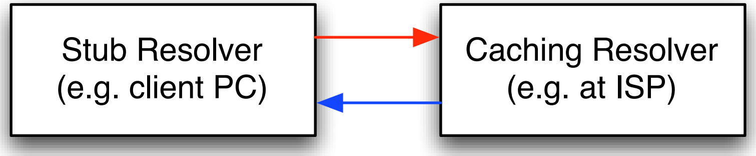
- A namespace
  - hierarchical, tree-like structure
  - labels separated by dots
- Nameservers
  - servers which respond to queries from clients, and make the data available
- Resolvers
  - clients which ask questions

www.uniforum.org.za



www.uniforum.org.za

- Answers which are already in the cache can be returned directly, with no recursive lookup required
- Items expire from the cache when they become stale



Root Servers

- Every recursive nameserver needs to know how to reach a root server
- Root servers are the well-known entry points to the entire distributed DNS database
- There are 13 root server addresses, located in different places, operated by different people
- The root zone is published by IANA

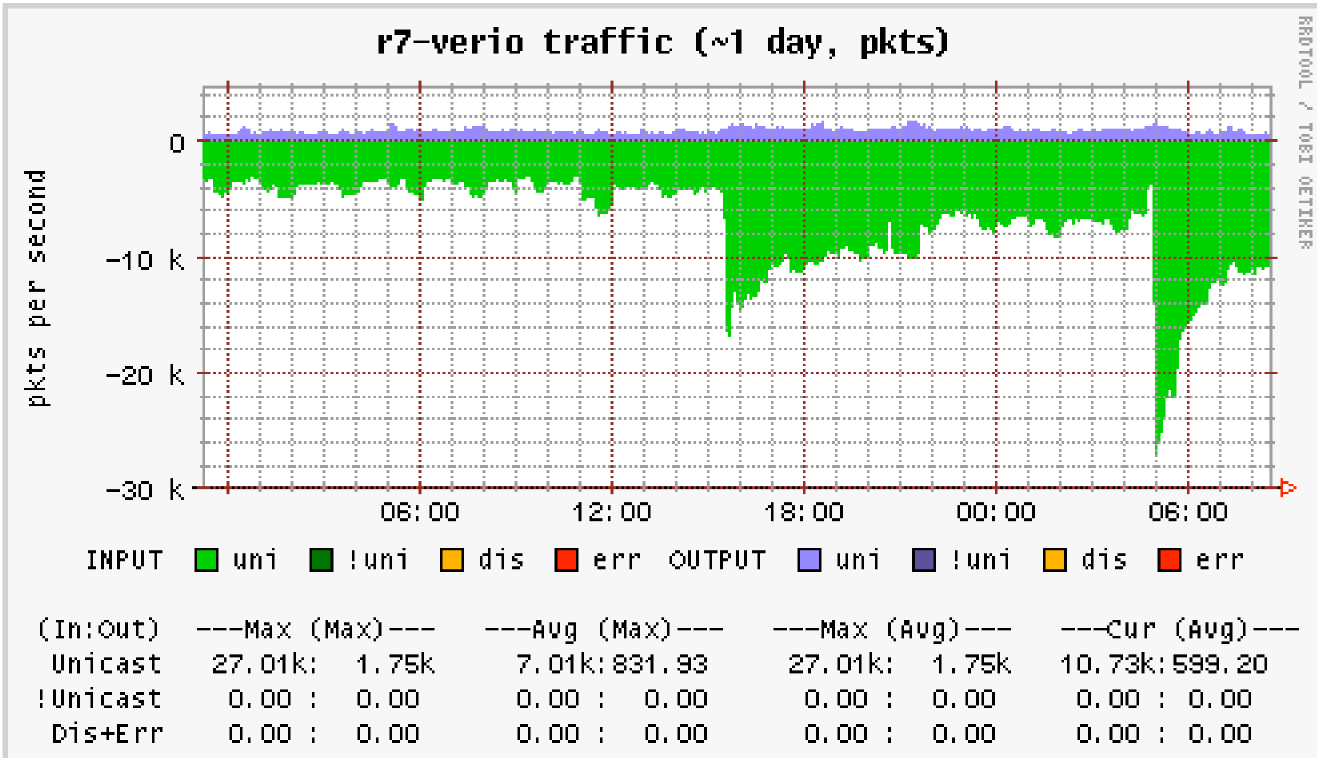
The Root Servers

A.ROOT-SERVERS.NET	Verisign Global Registry Services	Herndon, VA, US
B.ROOT-SERVERS.NET	Information Sciences Institute	Marina del Rey, CA,
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,
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

DNS Failure Modes

Challenges on the Root

- There have been a number of attacks on the root servers
- Distributed denial of service attacks can generate a lot of traffic, and make the root servers unreachable for many people
- Prolonged downtime would lead to widespread failure of the DNS



It's a Jungle Out There

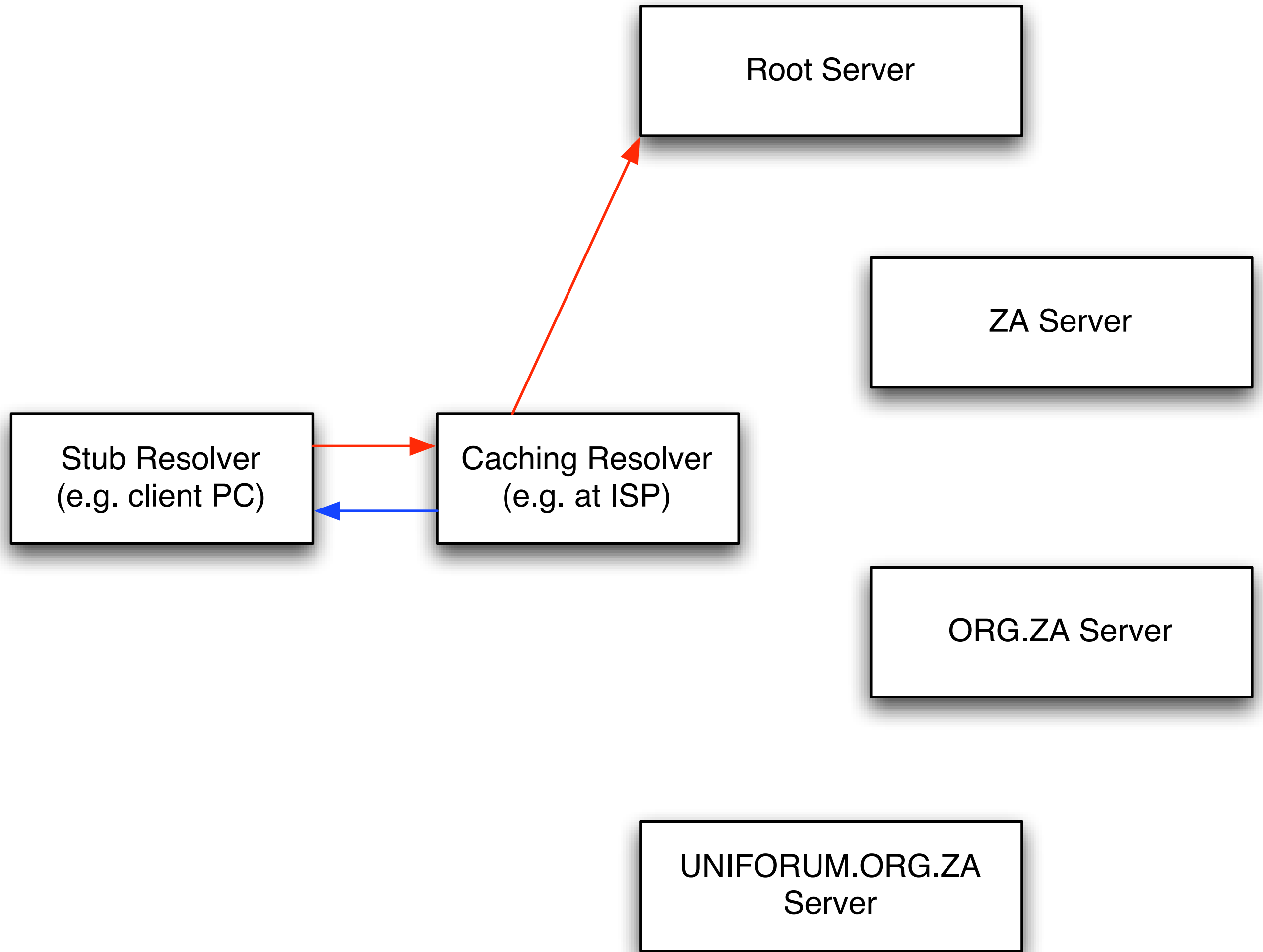
Global DNS Failure

- Probability of the entire DNS system failing is low
  - the most important data in the DNS (records which are frequently queried) are cached, usually with high(ish) TTLs
  - the individual root servers are run independently and are under substantial scrutiny
  - coordinated attacks on the root servers tend to be investigated vigorously

Regional DNS Failure

- If a region becomes partitioned from the Internet, or suffers a prolonged lack of access to the root nameservers for some other reason, the DNS may fail within that region
- Issues affecting small regions do not attract the same attention as issues affecting the whole network
- Regional DNS failure is much more likely than global failure

www.uniforum.org.za



Loss of Network

- Many countries depend on a relatively non-diverse set of external networks to reach the rest of the world
  - one under-sea cable; one satellite operator
  - a common circuit termination point in a telco hotel somewhere
  - an international network that is close to capacity, and which becomes useless if flooded with junk traffic

The Distributed F Root Nameserver

f.root-servers.net

- Has a single IPv4 address (192.5.5.241)
- Has a single IPv6 address (2001:500::1035)
- Requests sent to those addresses are routed to different nameservers, depending on where the request is made from
  - this behaviour is transparent to devices which send requests to F

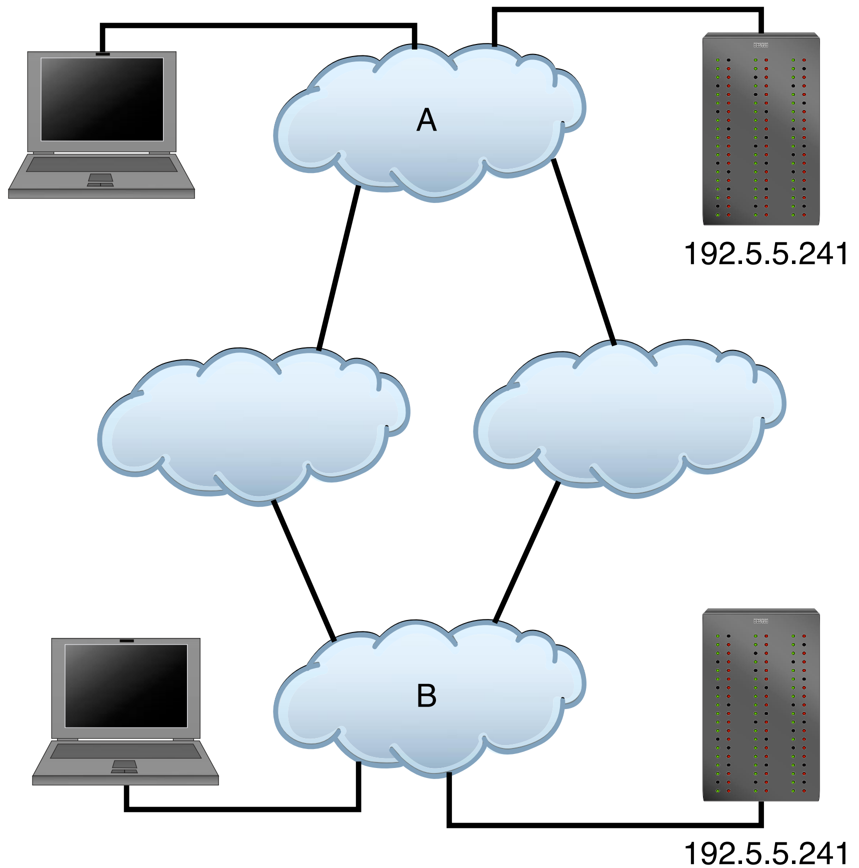
Unicast, Multicast

- Most traffic on the Internet is unicast
  - packets have a single destination
- Some traffic is multicast
  - packets are directed to multiple destinations

Anycast

- Traffic to f.root-servers.net is anycast
  - packets are directed to a single instance of F, but different queries (from different places) may land on different instances
  - anycast is identical to unicast from the perspective of the client sending a request

Anycast Routing



Hierarchical Anycast

- Some of the F root nameserver nodes provide service to the entire Internet (global nodes)
  - very large, well-connected, secure and over-engineered nodes
- Others provide service to a particular region (local nodes)
  - smaller

Hierarchical Anycast

- Each local node's routing is organised such that it should not, under normal circumstances, provide service for clients elsewhere in the world
- For more details, see:
    - `http://www.isc.org/tn/isc-tn-2003-1.html`

Failure Modes

- If a local node fails, queries to F are automatically routed to a global node
- If a global node fails, queries are automatically routed to another global node
- Catastrophic failure of all global nodes results in continued service by local nodes within their catchment areas

Failure Modes

- If a region loses international connectivity (e.g. an under-sea cable cut), access to the root nameserver is preserved by virtue of the region's local node
- since the root is reachable, other local nameservers are also reachable (e.g. ZA servers, ORG.ZA servers)
- since TLD servers are reachable, in-country traffic to locally-named services can proceed

Failure Modes

- A denial of service attack against F launched from outside the region is invisible to users within that region
- A denial of service attack against F launched from within the region is invisible to everybody else in the world
- A widely distributed denial of service attack will cause discomfort proportionate to the size of the region (probably, maybe)

Triangulation

- Many denial-of-service attacks use source-spoofed attack traffic
  - time consuming to track back through a network
  - attacks frequently stop before the trace completes
- Watching the relative reactions of local nodes to an attack can help identify the real source

Logistics and Administrivia

Sponsorship

- ISC is a non-profit company
- Equipment, colo, networks for remote nodes are paid for by a sponsor
- All equipment is operated exclusively by ISC engineers
- The sponsor covers the ISC's operational costs of running the remote node

Deployment Status

Global Nodes

- Palo Alto
- San Francisco

Local Nodes

- 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

Local Nodes

- 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**

The Nairobi F

Vital Statistics

- Physically colocated with the KIXP switch
- 100 Mbit/s connection to the KIXP
- Two redundant, much lower-capacity transit paths via two independent ISPs for management, measurement, zone transfers
- Cluster of two nameservers sharing the query load

Using the Local F

- You may be already using it
  - `traceroute f.root-servers.net`
  - `dig @f.root-servers.net hostname.bind chaos txt`

Before...

```
[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:~]$
```

... and After

```
[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