INTERNET SOFTWARE CONSORTIUM

f.root-servers.net

iWeek, September 2003

Joe Abley <jabley@isc.org>

The Basics

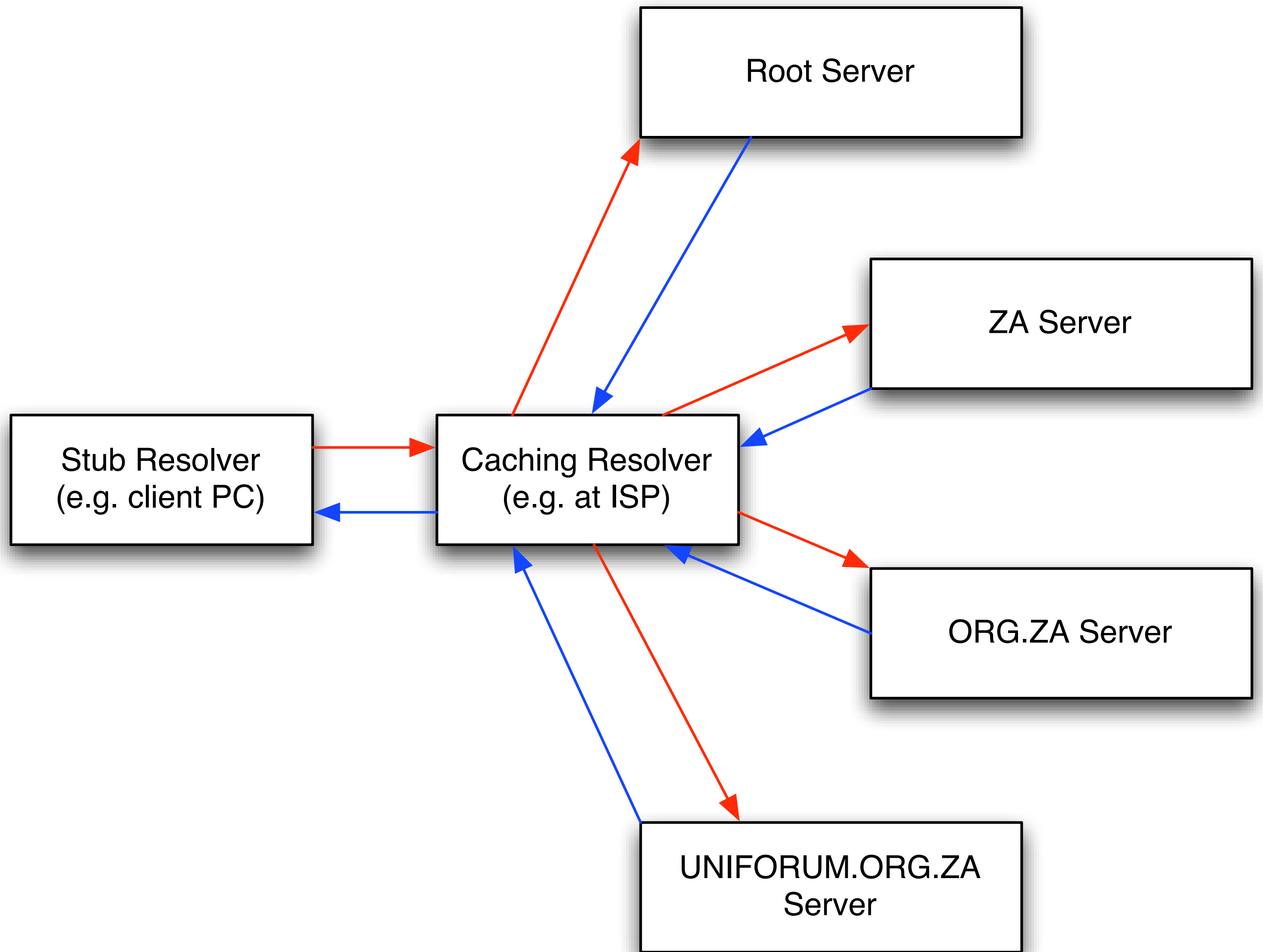
DNS

- 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

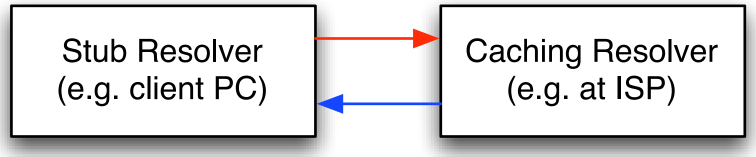
- A namespace
  - hierarchical, tree like structure
  - labels separated by dots
- Nameservers
  - servers which respond to queries from clients, and make the data available
- Resolvers
  - clients which ask questions

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

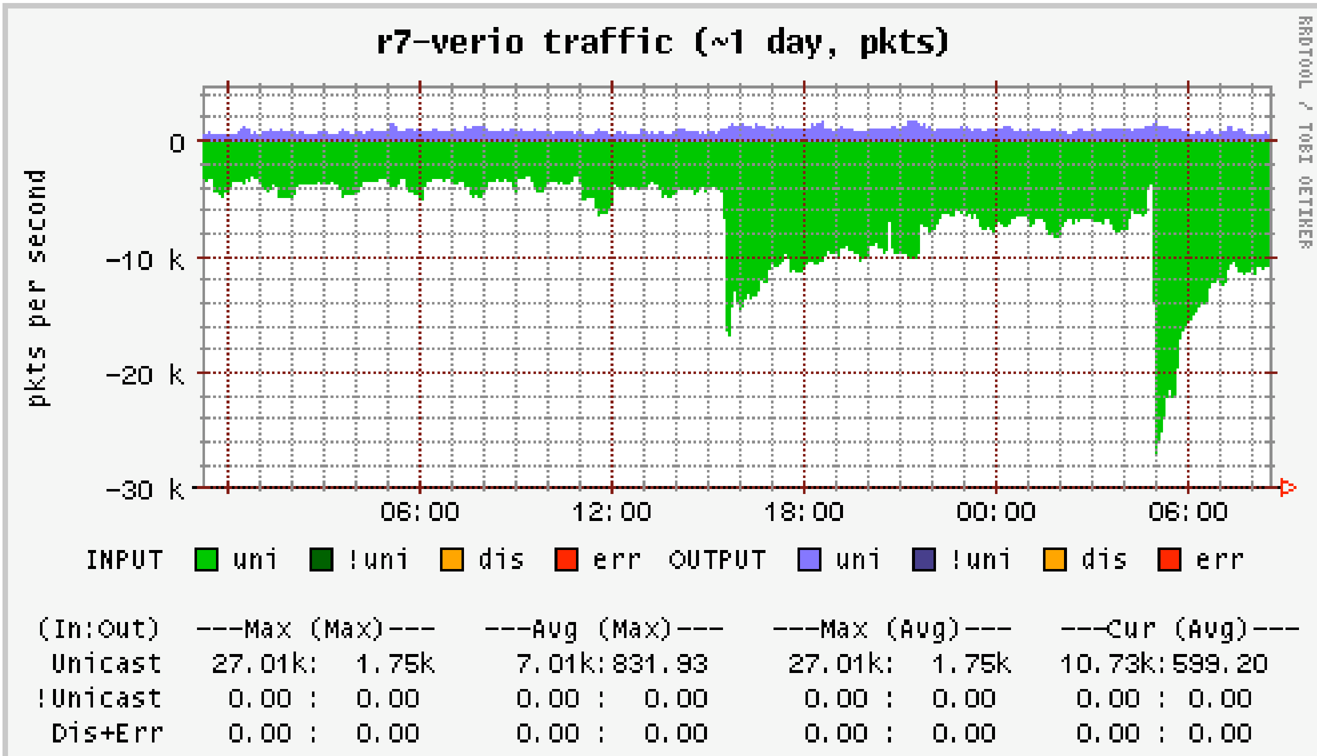
The Root Servers

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

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

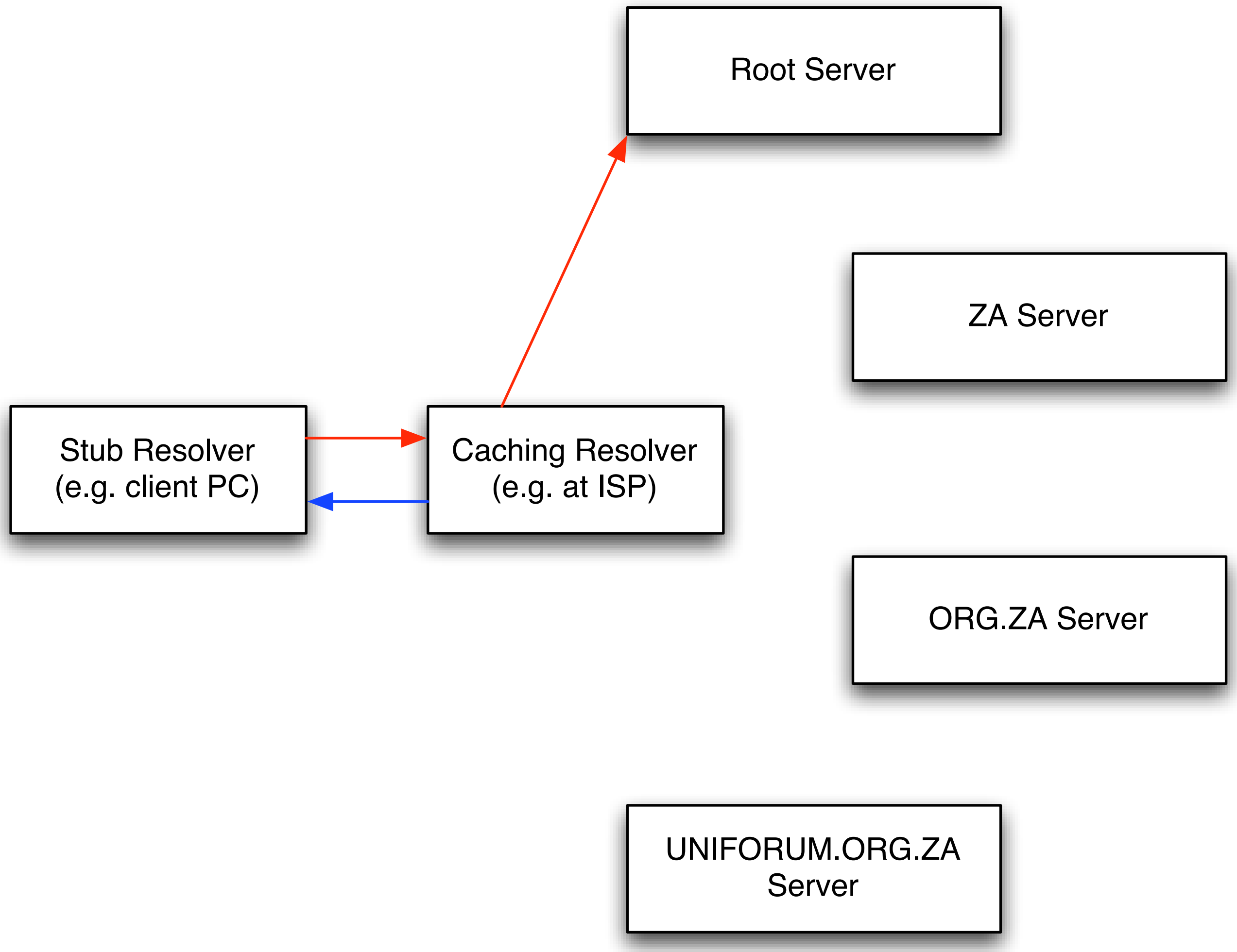
Global DNS Failure

- 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 partioned 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
  - 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

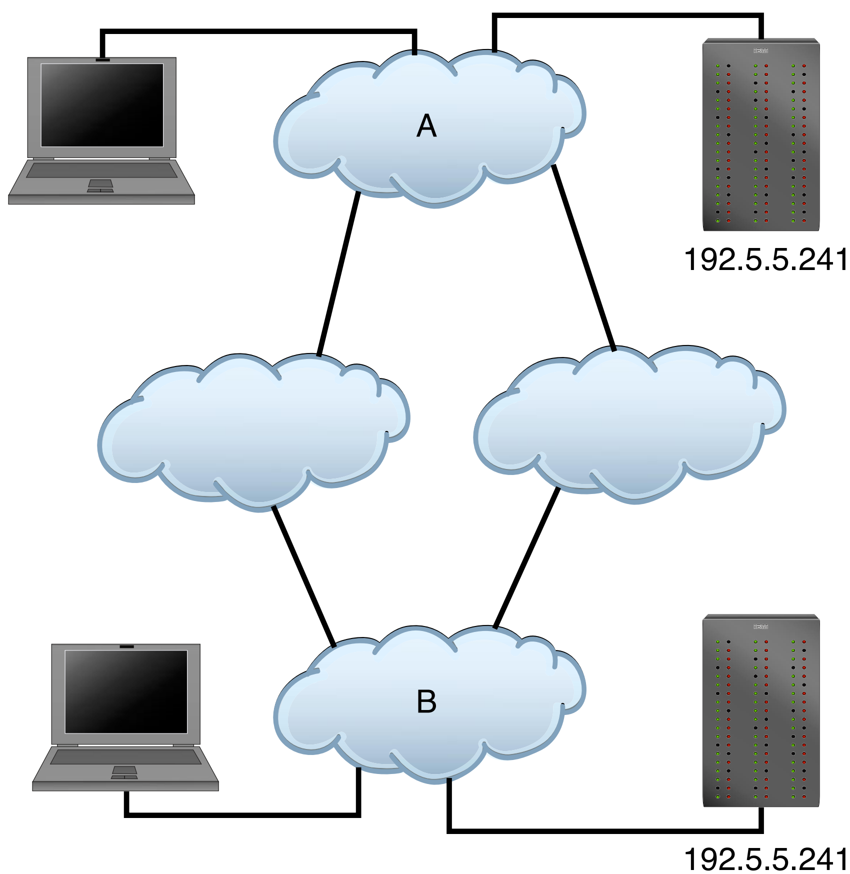
Unicast, Multicast

- 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

- Madrid, Rome
- São Paulo
- New York, Los Angeles, San Jose, Ottawa
- Hong Kong, Seoul, Beijing
- Auckland

Local Nodes

- Madrid, Rome
- São Paulo
- New York, Los Angeles, San Jose, Ottawa
- Hong Kong, Seoul, Beijing
- Auckland
- **Johannesburg**

Deployment Targets

- 10 local nodes live by the end of 2003
  - (we might need to revise that one)
- 20 more in 2004

The Johannesburg F

Vital Statistics

- Physically colocated with the JINX switch
- Dual 100 Mbit/s connections to the JINX
- 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`
- If you're not already using it, the way to get access is to peer with the F root node at the JINX
  - `http://www.isc.org/peering`

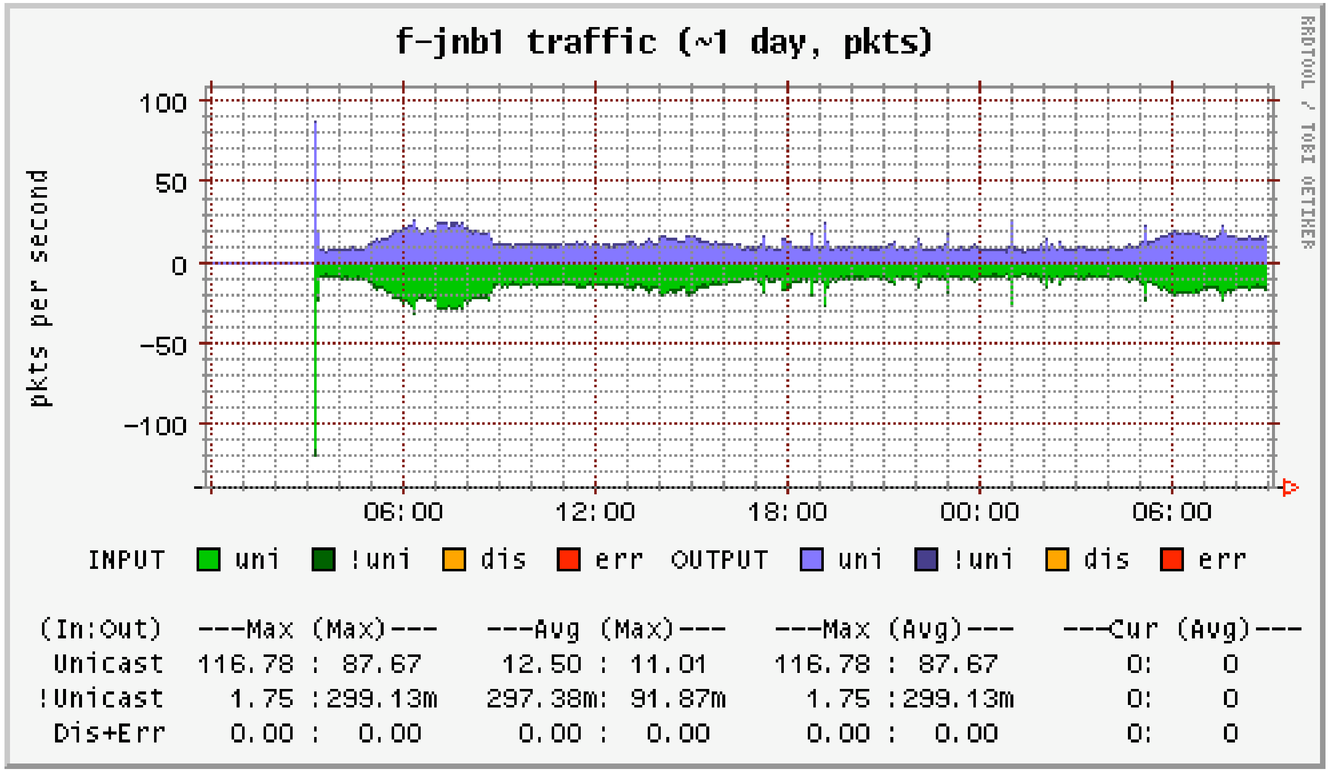
Before...

```
traceroute to f.root-servers.net (192.5.5.241), 30 hops max, 40 byte packets
 1  uunet-gw.barn.za.net (196.7.14.1)  6.488 ms  7.920 ms  0.571 ms
 2  router.barn.za.net (196.7.14.130)  55.080 ms  54.090 ms  39.162 ms
 3  s8-0-7chan23.gw1.cpt1.alter.net (196.31.167.105)  99.316 ms  136.754 ms  95.271 ms
 4  atm8-0-0sub100.ir2.mia16.alter.net (196.30.229.170)  309.513 ms  388.618 ms  322.437 ms
 5  POS0-1-0.IH4.MIA4.ALTER.NET (152.63.86.145)  307.761 ms  309.175 ms  289.307 ms
 6  202.at-5-1-0.XR2.MIA4.ALTER.NET (152.63.7.130)  249.434 ms  268.680 ms  323.183 ms
 7  0.so-4-2-0.XL2.MIA4.ALTER.NET (152.63.101.46)  370.243 ms  308.866 ms  290.180 ms
 8  0.so-3-0-0.TL2.ATL1.ALTER.NET (152.63.101.53)  349.110 ms  408.991 ms  335.088 ms
 9  0.so-7-0-0.TL2.SCL2.ALTER.NET (152.63.1.69)  333.937 ms  376.692 ms  491.727 ms
10  0.so-4-0-0.XL2.PAO1.ALTER.NET (152.63.54.82)  439.421 ms  418.440 ms  370.696 ms
11  POS1-0.XR2.PAO1.ALTER.NET (152.63.54.78)  418.243 ms  395.978 ms  374.415 ms
12  188.ATM9-0-0.BR1.PAO1.ALTER.NET (152.63.50.45)  396.263 ms  432.991 ms  433.469 ms
13  * * *
14  f.root-servers.net (192.5.5.241)  393.992 ms  373.653 ms  382.521 ms
```

... and After

```
traceroute to f.root-servers.net (192.5.5.241), 30 hops max, 40 byte packets
 1  uunet-gw.barn.za.net (196.7.14.1)  0.464 ms  0.413 ms  0.418 ms
 2  router.barn.za.net (196.7.14.130)  24.301 ms  29.350 ms  19.611 ms
 3  s8-0-7chan23.gw1.cpt1.alter.net (196.31.167.105)  59.583 ms  29.233 ms  80.713 ms
 4  fe1-0.br1.jnb7.alter.net (196.31.17.162)  99.377 ms  89.261 ms  58.475 ms
 5  198.32.142.14 (198.32.142.14)  60.405 ms  78.449 ms  94.946 ms
 6  f.root-servers.net (192.5.5.241)  68.080 ms  158.616 ms  109.683 ms
```

Day-One Traffic



Credits

- ISPA
- cisco Systems
- Uniforum South Africa
- Internet Solutions, UUNET South Africa
- Bucknet

INTERNET SOFTWARE CONSORTIUM

Questions

http://www.isc.org/misc/f-root-iweek-2003.pdf