INTERNET SYSTEMS CONSORTIUM

F root anycast: What, why and how

João Damas

ISC

Overview

- What is a root server? What is F?
- What is anycast?
- F root anycast. Why?
- How does ISC do it?

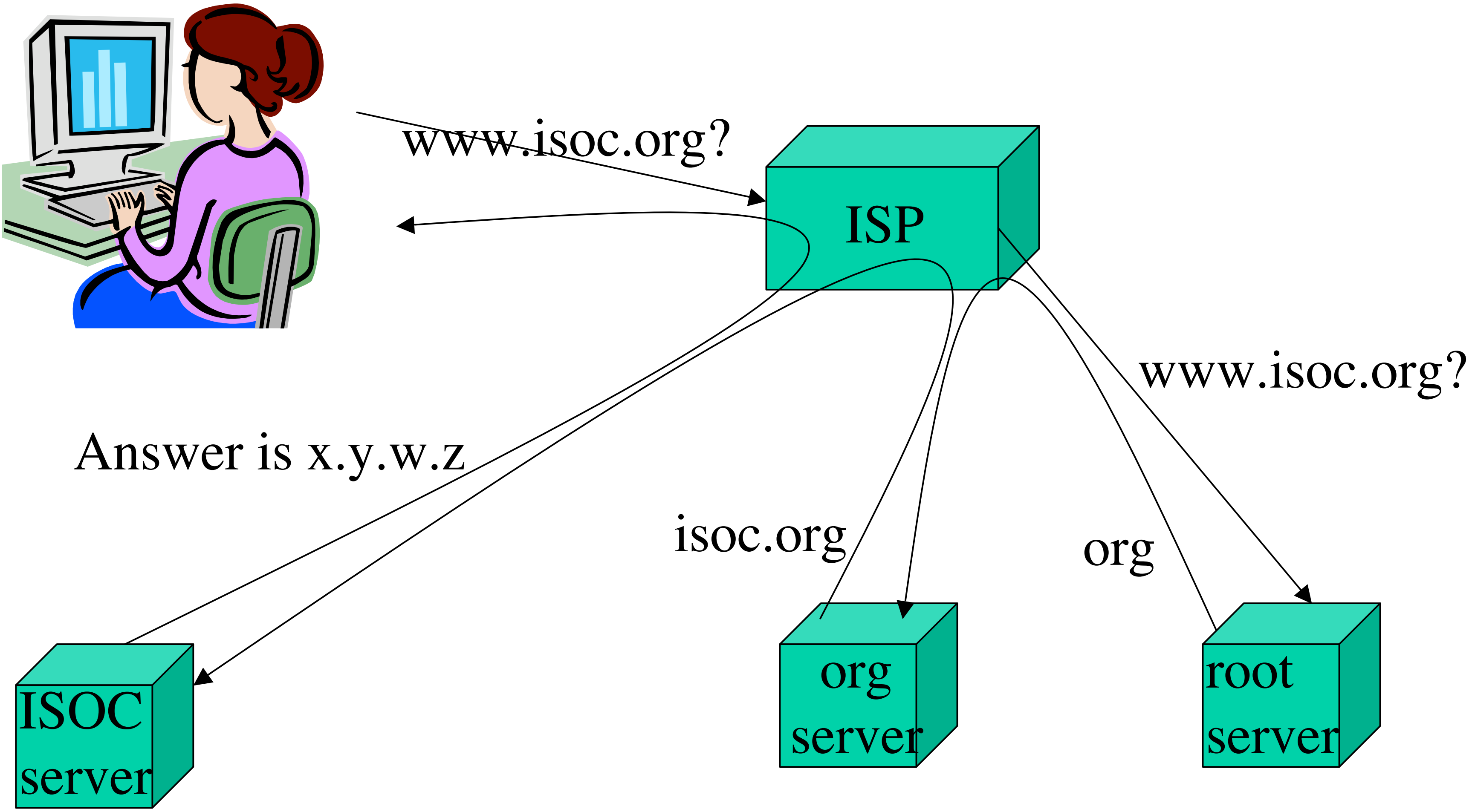
What is f.root-servers.net?

One the Internet’s official DNS root servers

The DNS

- –A tree-like lookup system
- –Converts human readable tokens into machine usable identifiers
- –Root servers are the entry point to the system
- –Caching is used throughout to avoid repetitive queries

DNS at a glance



The root zone

- The most important information provided by the root servers root servers is the root zone
  - .com -> list of servers
  - .net -> list of servers
  - .ch -> list of servers
  - .fr -> list of servers
  - .br -> list of servers
- Edited by IANA. Currently distributed by Verisign to all root servers

The root servers

- Serve the root zone
  - Provide the service itself
  - Currently limited to 13 distinct entries in the list
    - a.root-servers.net,…,m.root-servers.net
    - Purely technical role. Responsibility of each root server operator
  - A public service necessary for the correct functioning of the Internet
  - usually statically configured in resolvers

The root servers (ii)

- Root servers operators strive to make the service :
  - universal
  - Reliable and dependable
  - Quick

Challenges

- Internet is worldwide and getting more so every day
  - Root server distribution was not diverse enough in the network.
- Rise in traffic (attacks, misconfiguration, bad software)
  - Need to increase capacity

Impact of loss of service

- At a global level, the loss of a few root servers should not pose problems for some time (order of a few days) due to caching, as long as some root servers remain reachable
- At a regional level, loss of network reachability can have greater impact
  - Loss of upstream connectivity (eg, undersea cable cut or satellite link)
  - Caching information expires and even intra-region communication fails

Anycast

- Traditional traffic is unicast (one to one)
- The term anycast was created to define network services were multiple servers respond to the same IP address and provide the same service for that IP address (RFC 1546, Nov 1993)
- Works well for stateless protocols, such as UDP

Anycast (ii)

- The routing system decides where to send each packet, so that it reaches one (or more) destination that serves a given IP address.
- Two cases, according to which routing protocol is used
  - Localised
  - Internet-wide

Localised anycast

- A group of servers in the same subnet
  - Each server has the service IP address configured on a loopback interface and runs a routing daemon (ospfd) to announce itself as a router that can reach the service IP via its external network interfaces
  - Use OSPF ECMP to deterministically choose the destination server.
    - Uses the combination of srcIP, srcPort, dstIP, dstPort to generate a hash that will determine the next hop
  - Path selection is constant if the number of running servers does not change

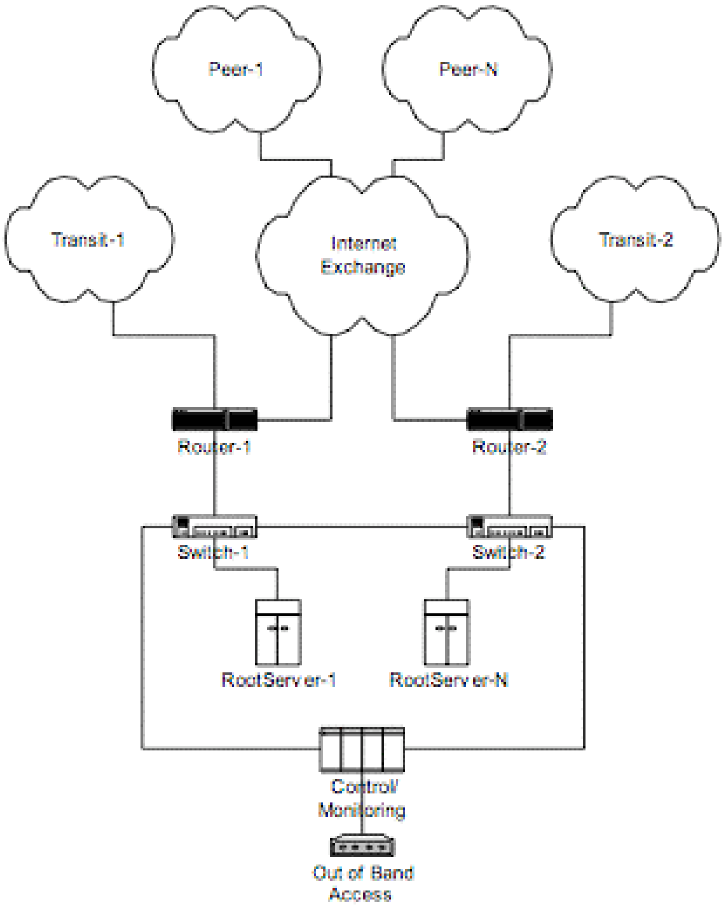
Internet-wide anycast

- Outside the local subnet.
- Internet interdomain routing is done with BGP
- A prefix/origin-AS combination is announced to peers from several points on the Internet.
- Similar to multihoming but the origin AS is composed of islands, there is no internal network communicating all the “exit” points.
- BGP selects “best” path at a given time. Subject to changes in path selection.

F root anycast setup

- http://www.isc.org/tn/isc-tn-2003-1.txt
- Uses localised and Internet-wide anycast
  - F root addresses
    - IPv4: 192.5.5.241
    - IPv6: 2001:500::1035
  - Each node is a cluster running ECMP ospf
  - Each node advertises the F root prefixes with origin AS 3557 to its BGP peers.

Typical node



Local and global nodes

- Global nodes
  - No restrictions on BGP announcements
  - Serve all potential hosts on the Internet
- Local nodes
  - Prefix announced with “NO-EXPORT” community in BGP. Have regional scope (usually)
  - Preferred locations are IXPs for their span of local ISPs

Failure modes

- When a local node fails it's BGP route is withdrawn and traffic flows to the next best path, usually a global node. This is to prevent cascading failures.
- When a local node is operating in a region that has become isolated, it keeps operating. Debating after how long the node should switch itself off to prevent stale information

Failure modes

- Two prefixes are announced via BGP, a /24 and a /23
  - /23 only from global nodes. Ensures there is a always a path to reach F root if the BGP choice for the /24 would result in no path
- Anycast also enables some forensics, by fractioning a potential DDoS attack into multiple ones. Monitoring of traffic is ongoing

Global nodes

- San Francisco
- Palo Alto

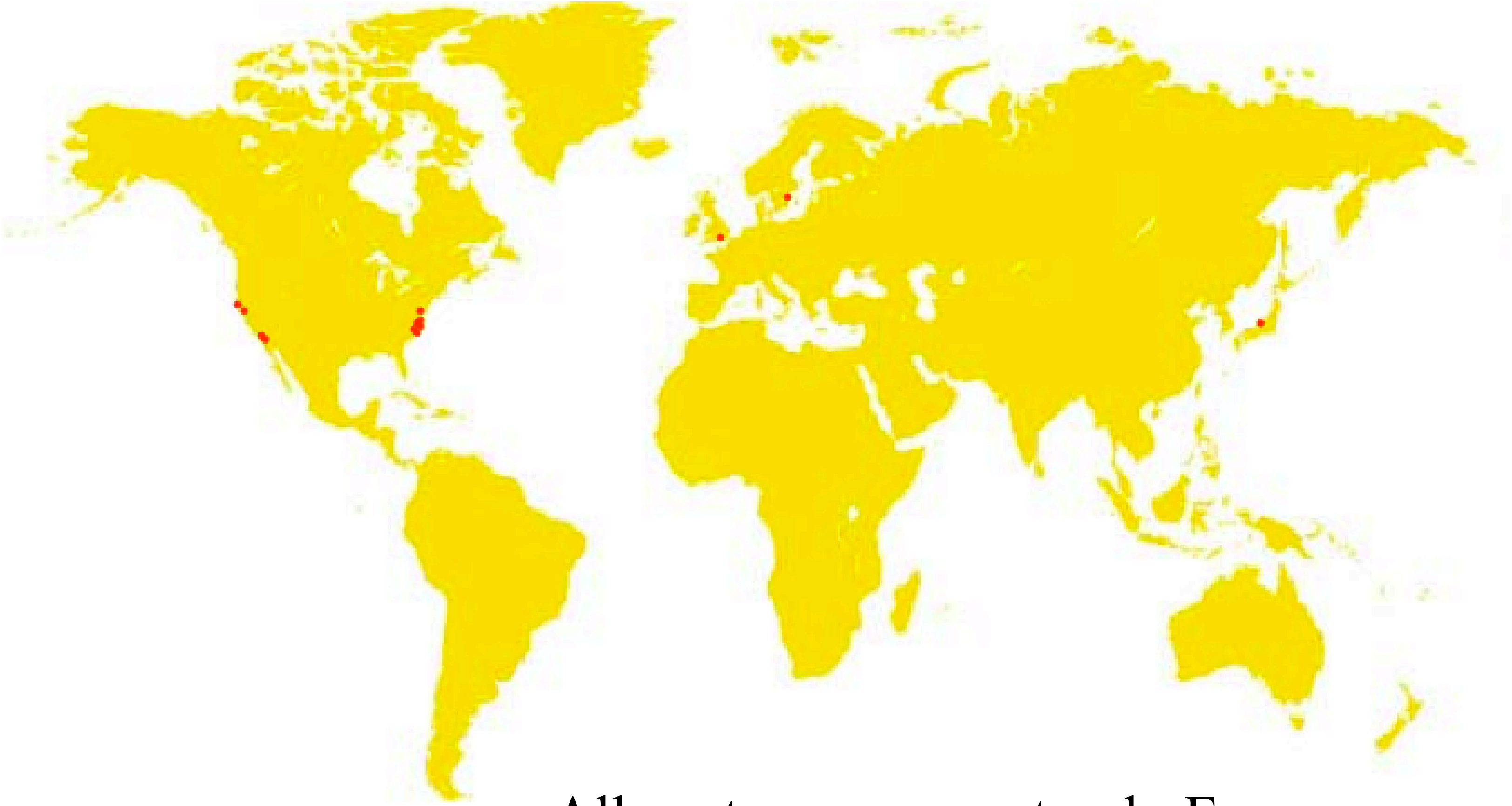
OSPF load balancing “anycast in a rack”

Unrestricted BGP

Local nodes

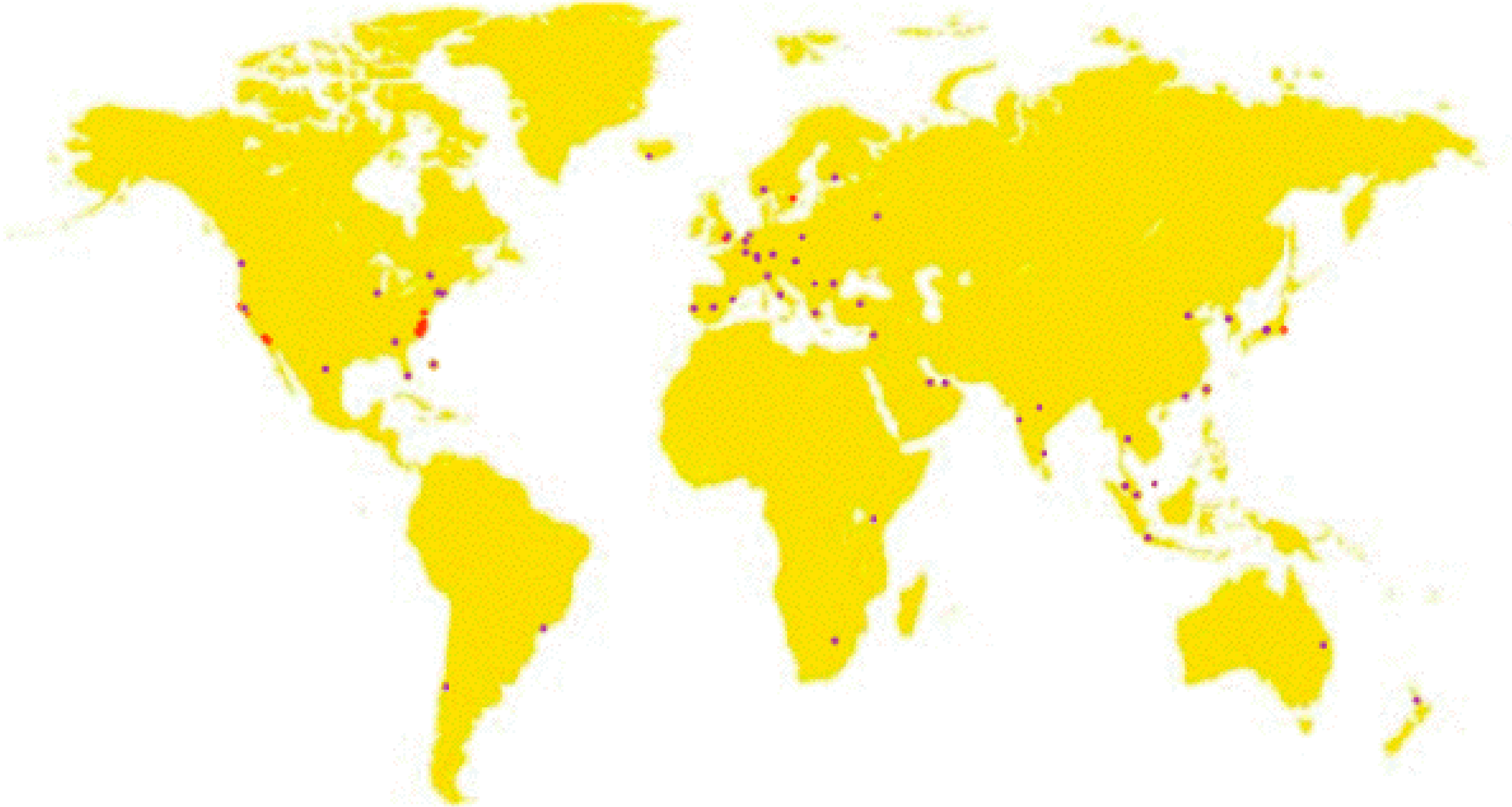
- Lisbon, Madrid, Barcelona, Paris, Amsterdam, Munich, Rome, Prague Moscow, London, Torino
- Tel Aviv, Dubai, South Africa, Kenya
- Beijing, Seoul, Osaka, Hong Kong, Taipei, Singapore, Jakarta, Chennai, Brisbane, Auckland, Dhaka, Karachi
- Monterrey, São Paulo, Santiago de Chile, Los Angeles, San Jose, New York, Toronto, Ottawa

Worldwide root server distribution -Before anycast



All root servers - not only F

Worldwide root server distribution -Now



All root servers - not only F

How ISC does this

- ISC is a not-for-profit organisation that works with multiple parties to publish Open Source software and carry out its operations
- Local sponsors and partners.
  - Always try to work with the local community
  - Enter an MoU that describes the interaction
  - Node administration is always the responsibility of ISC's engineers.
- ISC adapts to the local circumstances but always keeping service integrity first

Questions?

http://www.isc.org

Joao_damas@isc.org