// What information consumes is rather obvious:
it consumes the attention of its recipients.
Hence a wealth of information creates a poverty of attention,
and a need to allocate that attention efficiently among the
overabundance of information sources that might consume it. //
-- Nobel laureate Herbert Simon, 1995.

ucsd/sdsc/caida
bradley@caida.org
http://www.caida.org/outreach/presentations/

caida (cooperative association for internet data analysis)

research programs

- active: macroscopic topology project
- passive: (real–time) traffic workload characterization
- analysis of DNS root and gTLD server performance
- security issues, e.g., DOS measurement/analysis/modeling
- routing analysis and modeling
- bandwidth estimation methods and tools
- Internet Measurement Data Catalogue (IMDC)

other areas

- tool development
- new network visualization metaphors
- empirically guided Internet public policy
- outreach & education

visualization

Scientific visualization is the process by which some attributes or relationships are highlighted at the expense of others.

It is the selection of which attributes or relationships are to be preserved and which are to be hidden or distorted that drive the visualization task.

terminology

- IP – Internet Protocol
  - protocol used to connect different IP–based data networks together
  - IP address is fundamental
  - this presentation drops the word "address"
- AS – Autonomous System
  - collection of contiguous IPs which can announce IP prefixes
  - AS number uniquely identifies an AS (ISPs may have more then one)
  - this presentation drops the word "number"
- BGP – Border Gateway Protocal
  - primary protocol used by ASes to exchange routes
  - BGP table is a collection of BGP routes seen at a single router
- RTT – Round Trip Time
  - time to send and receive a response between two IP addresses
- protocol
  - collection of rules that define a service (HTTP, TCP, etc)
  - defined by port address

analysis of Domain Name System (DNS)

DNS is an indispensable Internet component

- two sets of 13 machines = root & gTLD name servers
- CAIDA has done several studies of multiple dimensions of DNS performance
  - passive monitoring of performance
  - active probing of latencies
  - analysis of root name server logs
- the infrastructure is changing now...
  - use of multicast, anycast

visualization

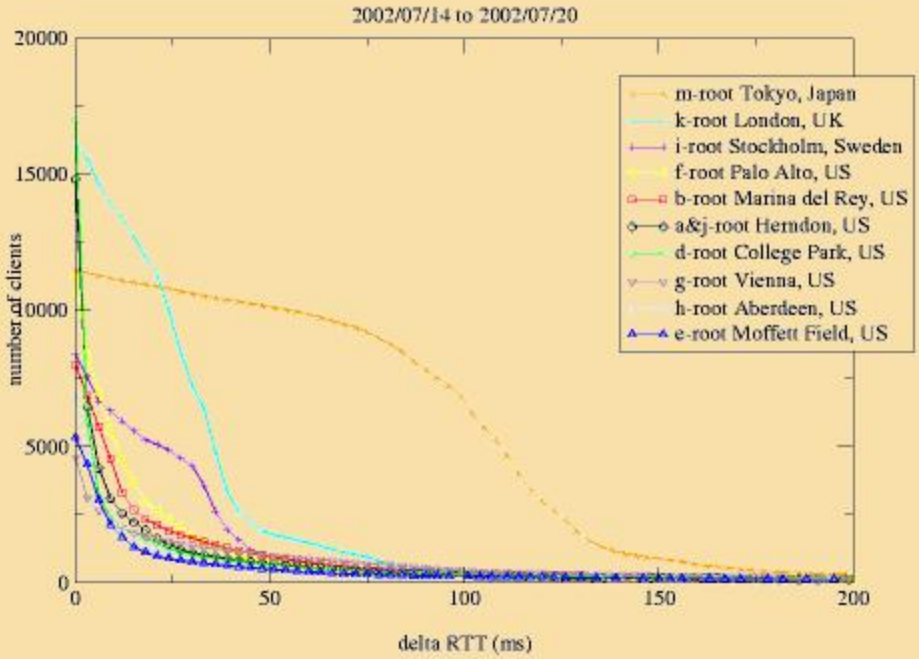
- DNS server root groups
- effect of removing a DNS root server
- effect of moving a DNS root server

DNS server root groups

	GROUP 1 EUROPE			GROUP 2 US-East				GROUP 3 CA US-West			GROUP 4 Tokyo Japan
	k-root	i-root	k-peer	a-root	g-root	h-root	d-root	f-root	e-root	b-root	m-root
k-root	0	89	105	124	127	125	125	146	138	162	207
i-root	89	0	130	144	158	152	155	174	164	188	218
k-peer	105	130	0	173	164	169	167	186	178	198	233
a-root	124	144	173	0	68	67	64	109	105	129	209
g-root	127	158	164	68	0	64	61	108	103	121	202
h-root	125	152	169	67	64	0	59	94	89	119	195
d-root	125	155	167	64	61	59	0	101	95	120	202
f-root	146	174	186	109	108	94	101	0	65	86	170
e-root	138	164	178	105	103	89	95	65	0	90	165
b-root	162	188	198	129	121	119	120	86	90	0	189
m-root	207	218	233	209	202	195	202	170	165	189	0

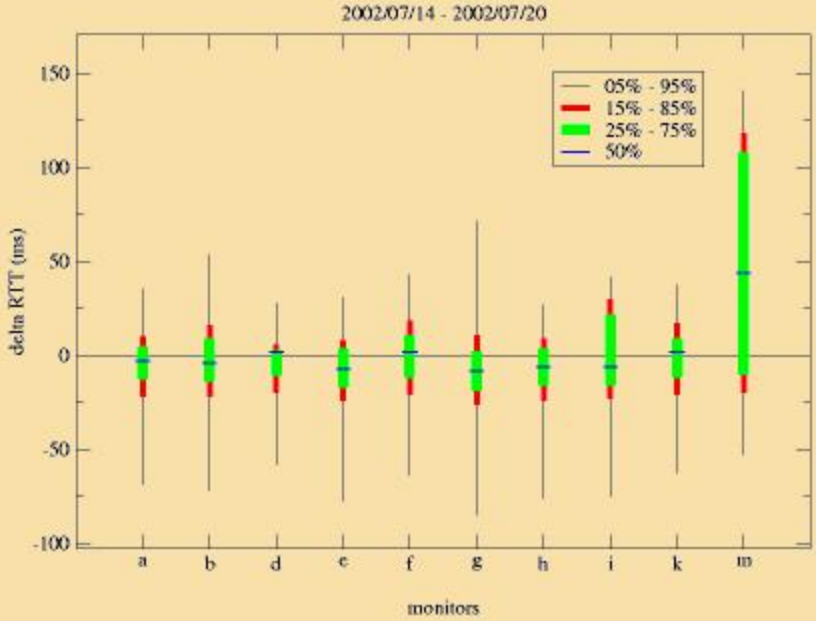
- presentation
  - roots grouped by "RTT client difference"
- axis
  - color – average absolute difference between clients' RTT
  - X – group and root server
  - Y – root server
- highlights
  - color helps the eye to cluster values into groups

effect of removing a DNS root



- presentation
  - number of clients whose delta RTT was greater, when the DNS root was removed, than the RTT on the X–axis
- axis
  - X – delta RTT

effect of moving a DNS root



- presentation
  - distribution of clients' delta RTT
- axis
  - color – shows the top 90%, 75%, 50%, 25%, and 5% points of the distribution
  - X – individual DNS root servers
  - Y – delta RTT
- highlights

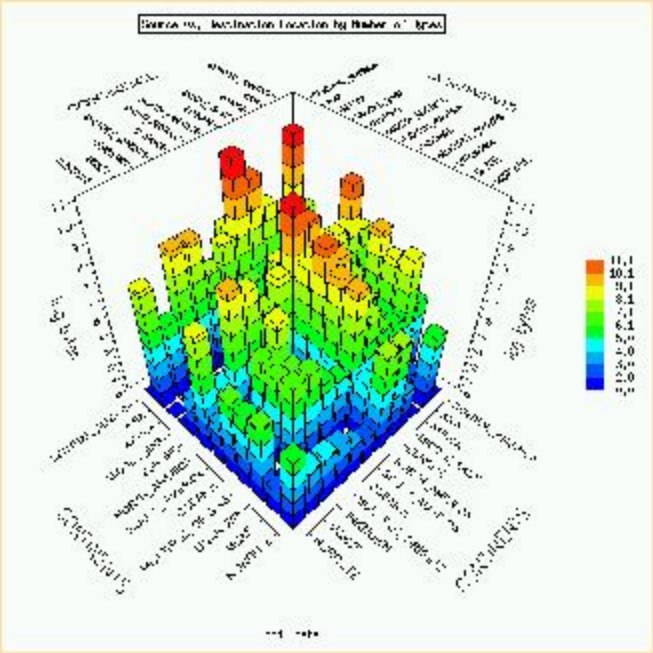
passive measurements of Internet data streams

- develop techniques of high speed traffic sampling
- measure busy OC3, OC12 and OC48 backbone links
  - CAIDA has the only publically available OC48 backbone IP flow data in the world
    - note: sprintlabs has private oc48 data
- study how user activities produce torrents of bytes
  - testing TCP models in presence of bursty cross traffic
  - detection of long running streams
  - tracking Internet usage patterns
- IETF WG developing flow measurement standards
  - Nevil Brownlee (CAIDA/U.Auckland) and David Plonka (U.Wisc) co-chairs

visualization (next slides)

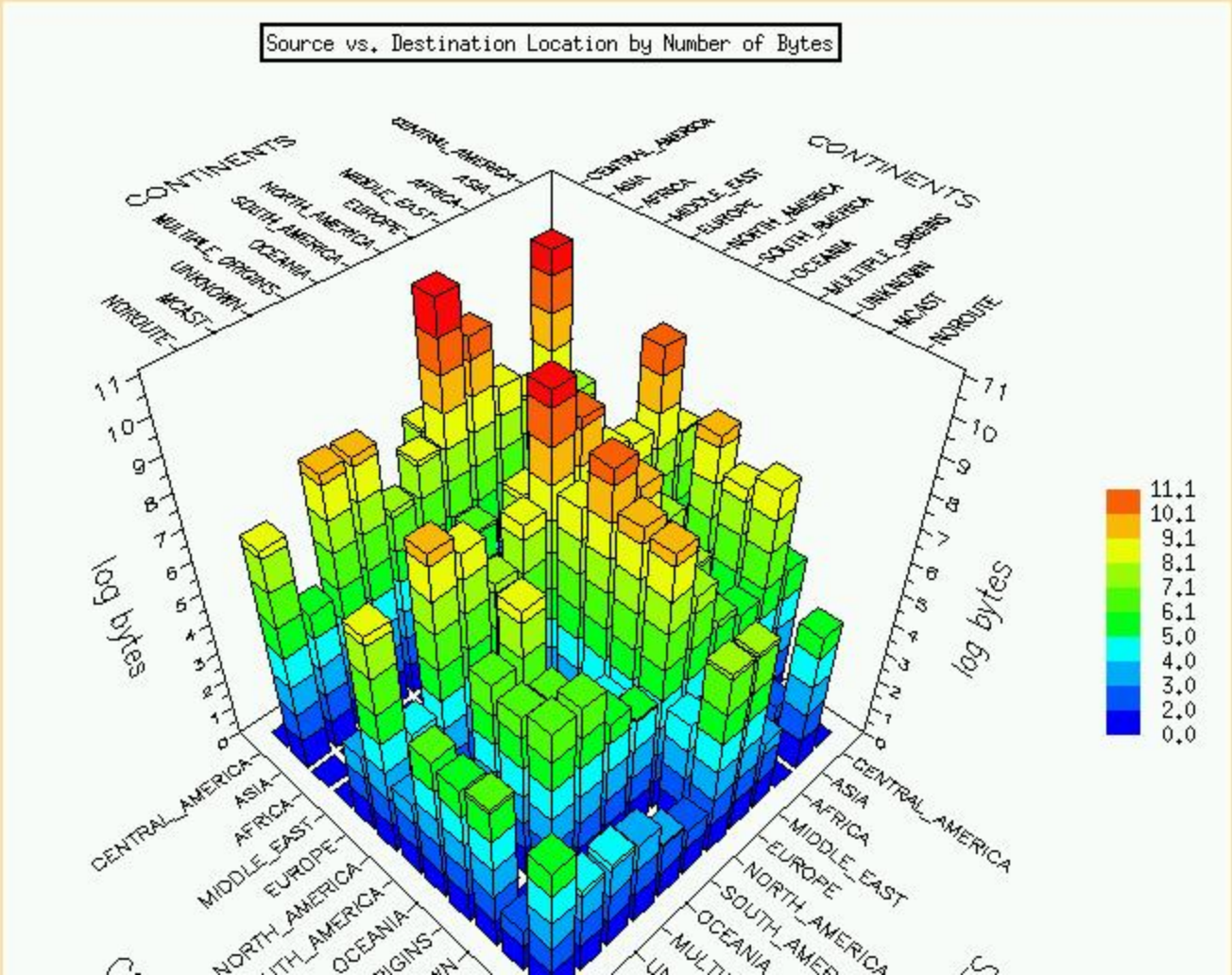
- continent to continent matrix
- breakdown by protocol

continent to continent matrix

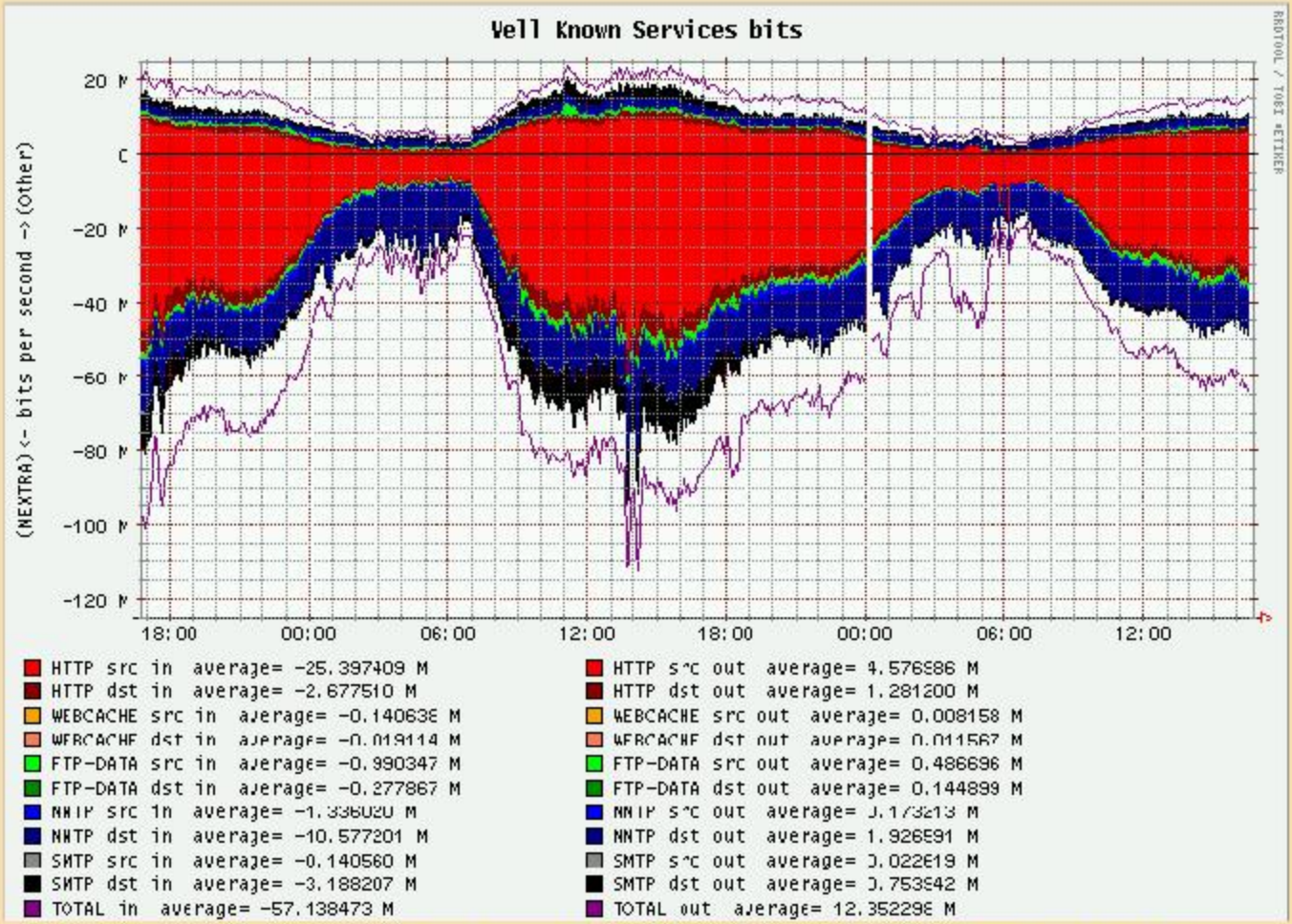


- presentation
  - traffic aggregated by country of announcing AS's headquarters
- axis
  - color – protocol of traffic
  - left – source continent (sorted by total bits sent)
  - right – destination continent (sorted by total bits sent)
  - z – logarithm of the number of bytes sent
- highlights
  - continents with low and high bit rates

continent to continent matrix



traffic breakdown by protocol



- presentation: flowscan tool -- flexibly shows traffic by protocol

interdomain routing

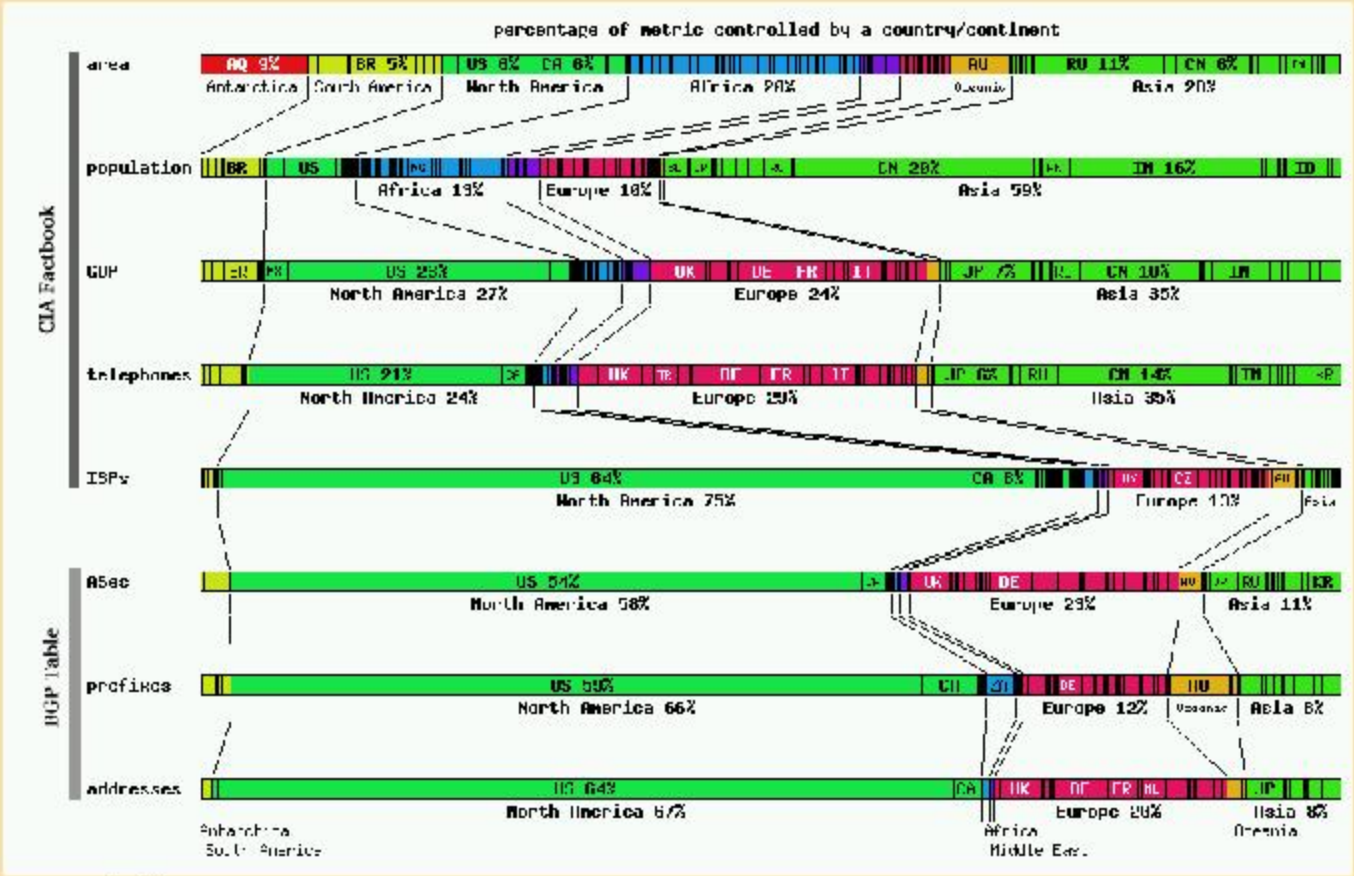
routing dynamics analysis

- evolution of global routing system
  - { AS, prefix, IP address } level granularities
- prerequisite to improvement of routing infrastructure
  - robustness, performance, integrity
- analyzing two–hour snapshots of BGP tables

visualization

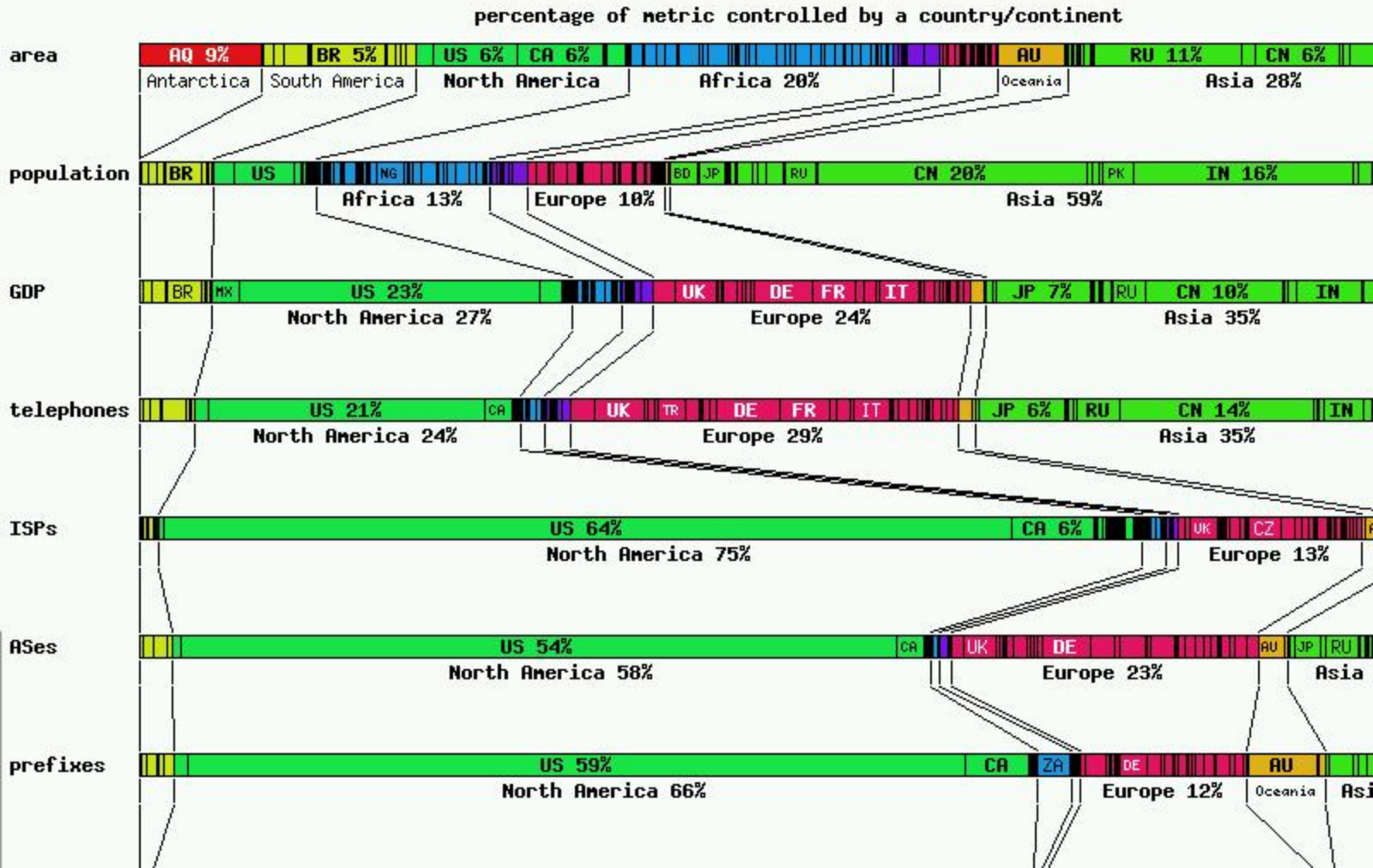
- geopolitical distribution of Internet resources

geopolitical distribution of Internet resources



- presentation
  - geographic breakdown of Internet space by headquarters of announcing AS
  - color – country
  - CIA Factbook – presented in the online CIA Factbook

geopolitical distribution of Internet resources



macroscopic topology project

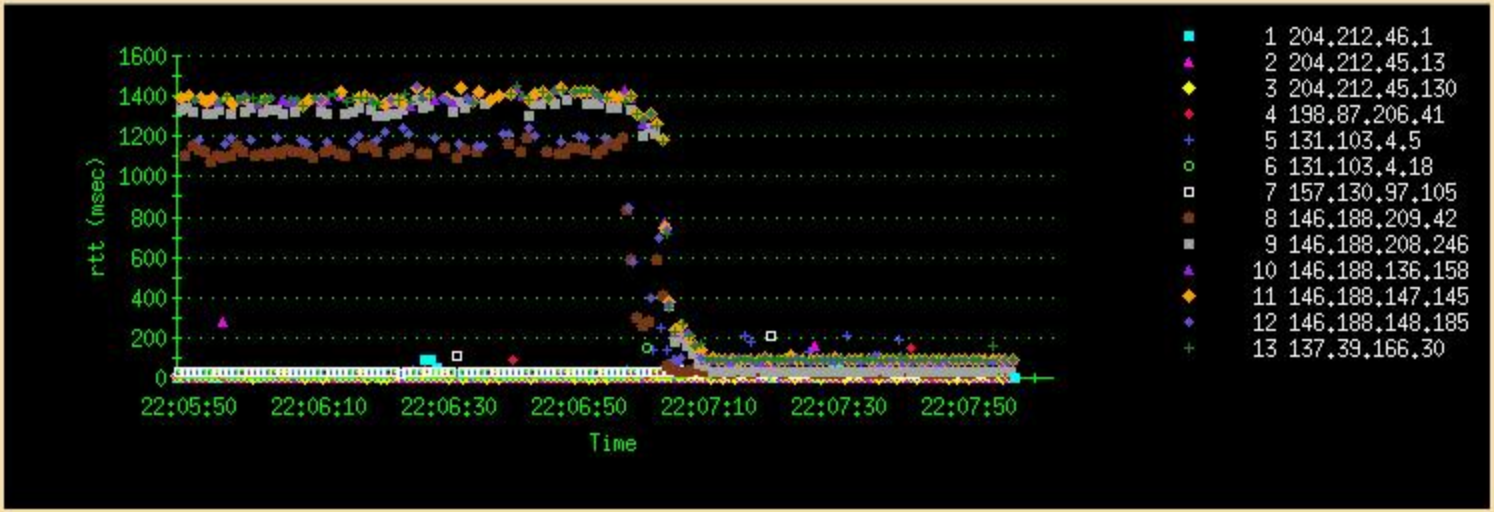
connectivity analysis

- massive macroscopic traceroute data – largest IP topology data set in world
  - funding from NSF, DARPA, commercial sources
- establishing framework for topology analysis
  - combinatorial structures (trees, acyclic part, core, giant component)
  - description by Weibull distributions
- mapping IP –> AS –> city, country –> latitude, longitude
  - BGP routing tables: IP –> prefix –> AS
  - whois registry/hostnames: AS, IP –> city,country –> latitude,longitude
  - note: there is no public trusted source of this data at this time (even CAIDA's has not been funded in 2 years)

visualization

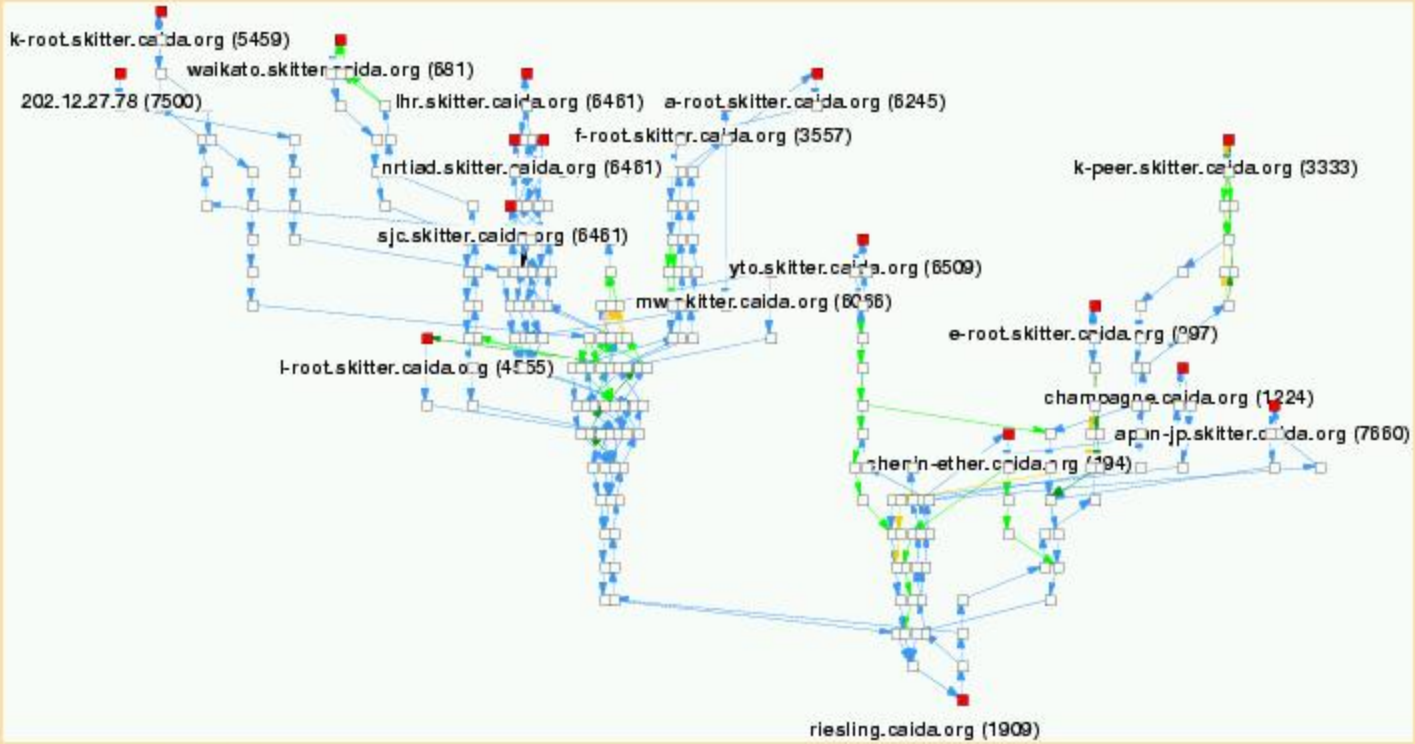
- IP paths: single source, single destination
- IP paths: single source, few destinations
- AS/country paths: single source, many destinations
- IP paths RTT: single source, many destinations
- IP topology: many sources and destinations
- AS topology: many sources and destinations
- geopolitical distribution of Internet resources

IP paths single source, single destination



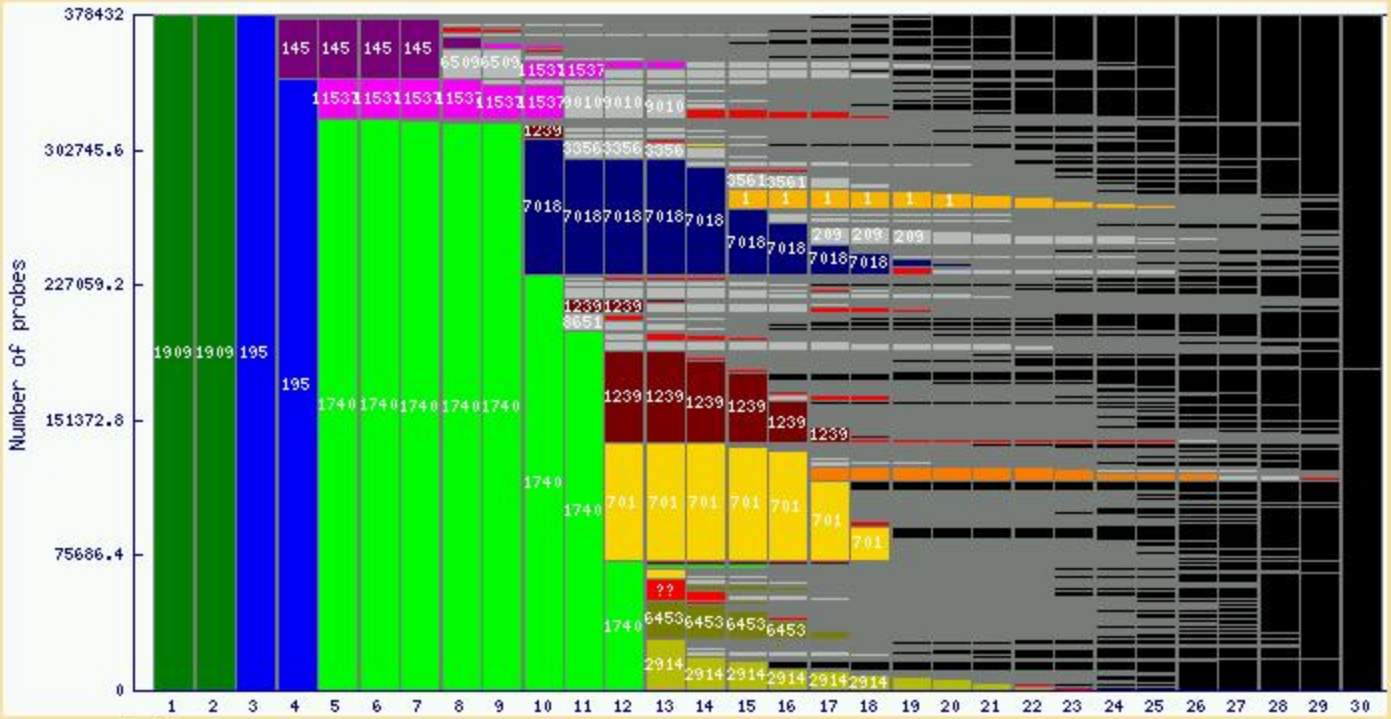
- presentation
  - IP path and intermediate RTT changes over time
- axis
  - X – time in minutes
  - Y – RTT in milliseconds
- highlights
  - path changes
  - points of large latency

IP paths: single source, few destinations



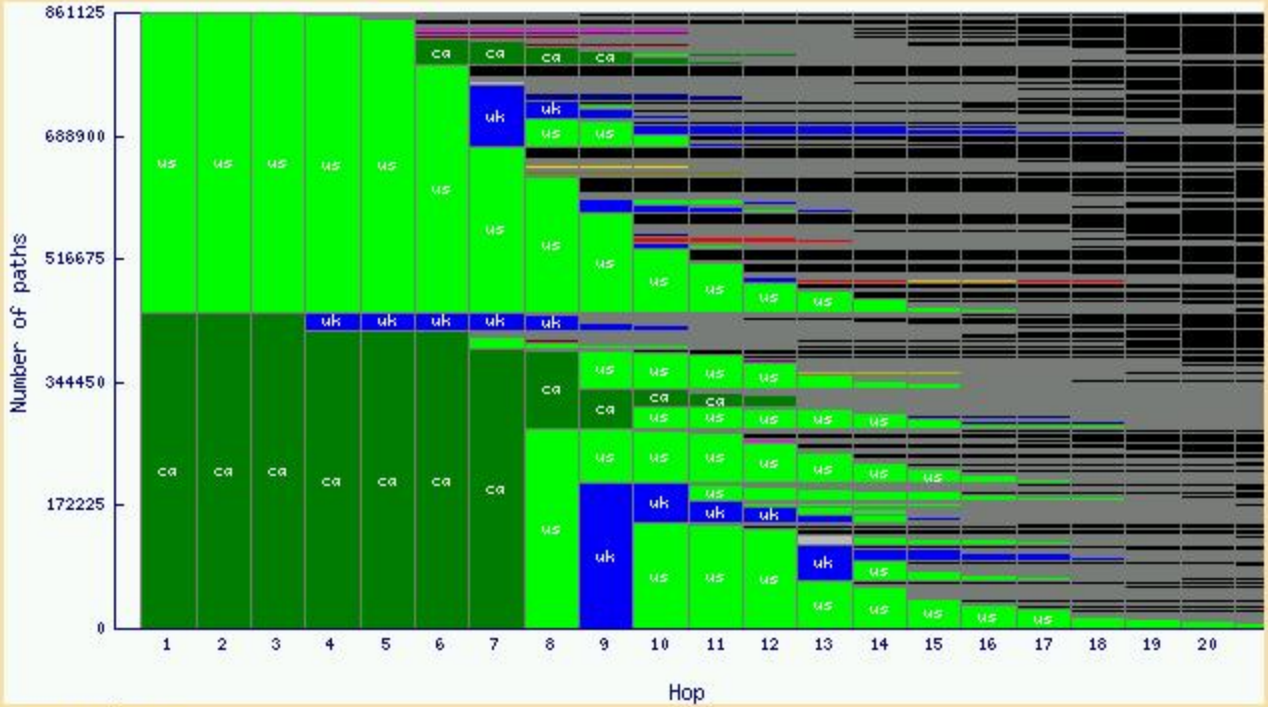
- presentation
  - bidirectional IP path between single source and few destinations
- axis
  - X – AS that announces the IP address
  - Y – hop count from source
- highlights

AS paths: single source, many destinations



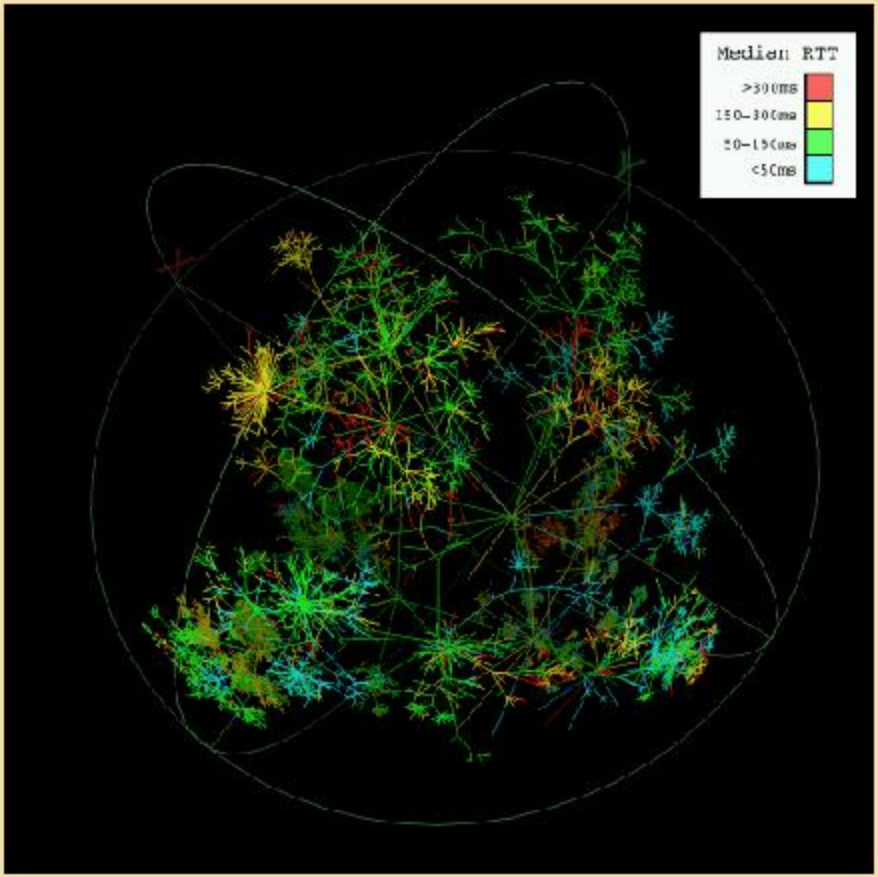
- presentation
  - proportion of traces passing through a given AS path from a single source
- axis
  - color – AS of a given hop.
    - black means the path has ended
    - grey is multiple tiny paths
  - X – distance in terms of the number of IP hop count
  - Y – number of paths which followed from the previous hop

country paths: two sources, many destinations



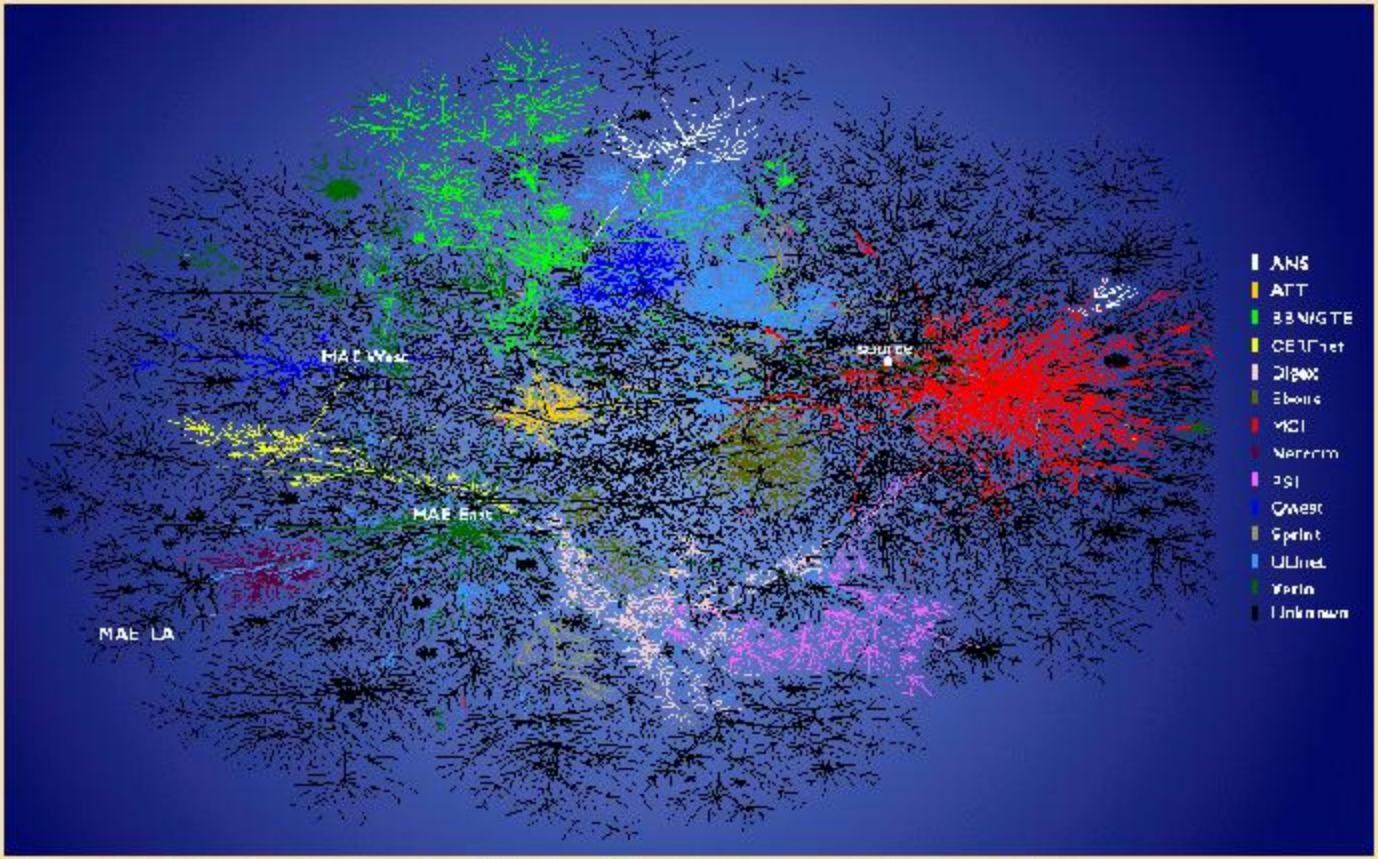
- presentation
  - proportion of traces passing through a given country path from two sources
- axis
  - color – country of a given hop.
    - black means the path has ended
    - grey is multiple tiny paths
  - X – distance in terms of the number of IP hop count
  - Y – number of paths that followed from the previous hop

IP paths RTT: single source, many destinations



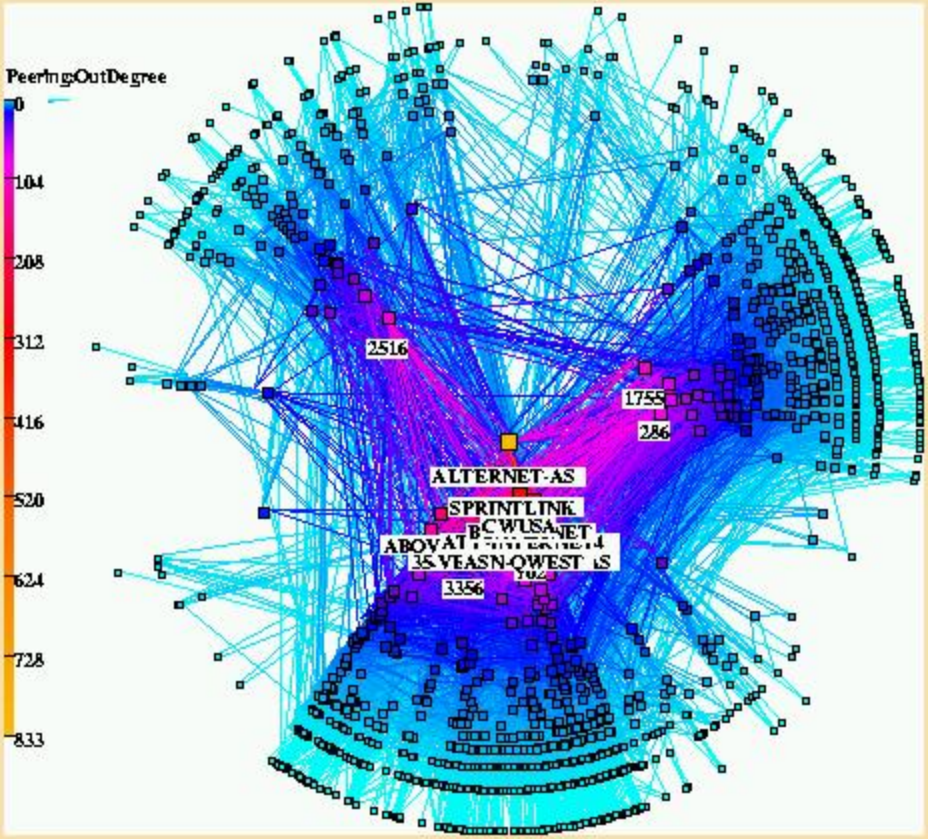
- presentation
  - shows the maximum latency experienced for destination along a given IP path
- axis

IP topology: many sources, many destinations



- presentation (note –– bell labs / lumeta layout algorithm)
  - IP spanning tree (does not include all links)
- axis
  - color – AS of longest matching prefix

AS topology many sources, many destinations



- presentation
  - IP forward path –> AS topology
- axis
  - color,radius – logarithmic of AS outdegree

security issues

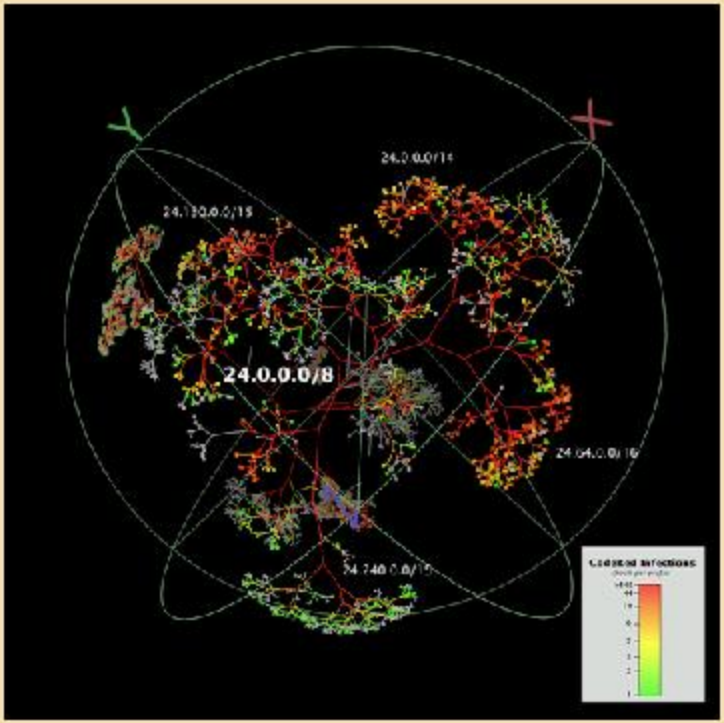
global denial of service activity

- backscatter methodology (CAIDA invented)
  - detecting denial–of–service (DOS) activity on the global Internet
- monitoring spread of worms in the networks
  - Nimda, Code Red, Sapphire, ... (to be continued)
- first (and remains primary) publically available data quantifying DOS

visualization

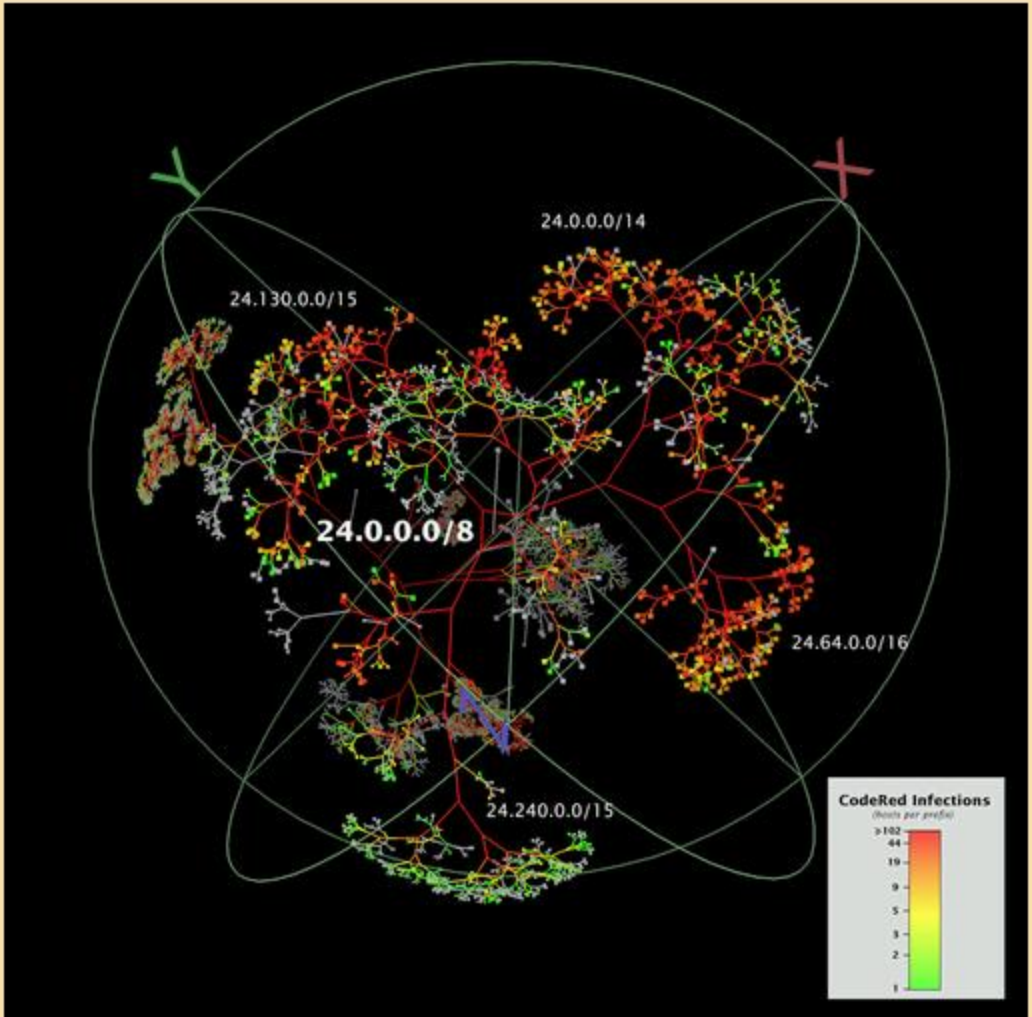
- number of Code Red infected hosts in 24.0.0.0/8

number of Code Red infected hosts in 24.0.0.0/8



- presentation
  - number of infected hosts in BGP prefixes
- axis
  - node – BGP routed prefix, or less specific prefix
  - link – connects more and less specific prefixes
  - color – number of infected hosts
  - X,Y – 3D hyperbolic geometry (fisheye distortions)

Number of Code Red infected hosts in 24.0.0.0/8



summary: visualization efforts at caida

remark: visualization is a medium, not an end

- visualization is tool to increase our ability to think
- visualization can help make us smart
  - it can also make us stupid by misadvised mappings and unworkable user interfaces

CAIDA's focus is on analysis rather than visualization

- however, we recognize that the community looks to CAIDA for visualization results
  - research and operational communities
  - since we have respected experience with analysis

Bradley Huffaker
ucsd/sdsc/caida
bradley@caida.org

http://www.caida.org/outreach/presentations/