



state of the art in visualization of Internet data:
topology, workload, performance and routing

11 January 2000

*scientific apparatus offers a window to knowledge,
but as they grow more elaborate,
scientists spend ever more time washing the windows.
-- Isaac Asimov*

ke claffy, UCSD/SDSC/CAIDA
ke@caida.org
www.caida.org

SDSC visualization research

- SDSC Vislab research:
 - modeling complex conditions
 - extending interactions
 - scalable visualization toolsets

- CAIDA research at SDSC:
 - analyzing characterization criteria
 - correlating different datasets
 - visualizing events and trends

CAIDA vis: aiding Internet analyses

identify and present insights into

- topology
- historical workload trends
- active performance evaluation
- routing efficiency

represent complex behavior, including

- correlation of large datasets
- enabling drill-down

CAIDA vis: Internet visualization

- helping to manage explosive growth throughout the Internet
- enhancing ability for providers and users to interpret vast quantities of data in real-time

@@@

- creating user-friendly integration with network utilities and control systems

CAIDA vis: challenges

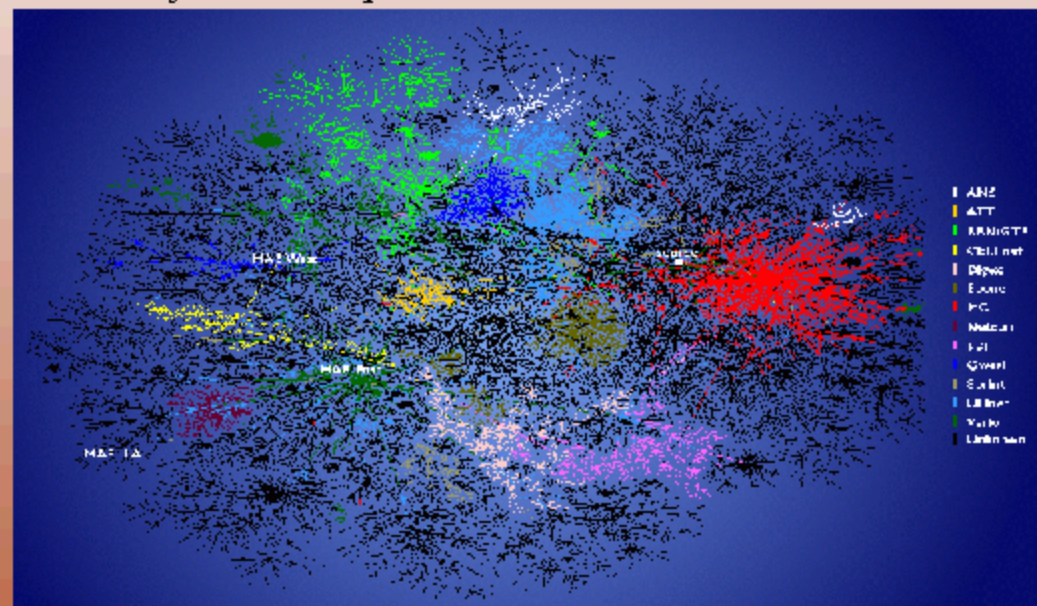
- manage complexity
 - enable inter- and intra-ISP analysis and feature detection
- develop new methods for data collection, reduction, aggregation, and mining (GByte or Tbyte datasets)
- manage data that is geographically and logically distributed as well as dynamically changing

=> visualizing topology: overview

- several methods available
 - geographical
 - logical
 - hyperbolic (T. Munzner, Stanford)

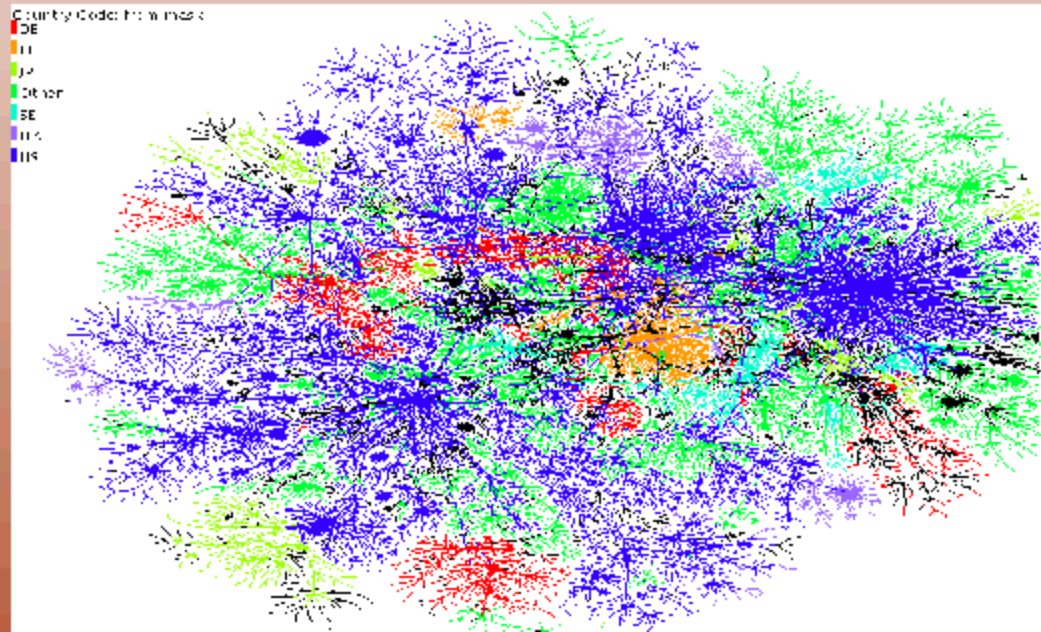
visualizing topology: skitter

- macroscopic, infrastructure-wide
- dynamically discover/depict topology (& b/w)
- correlate path perf. w events, e.g. BGP
- identify critical pieces of infrastructure



skitter: infrastructure-wide measurements

- 17 monitors (inc. 1 root name server)
- multiple dst lists (29k servers, 36k dns)
- architecture:
 - parallel ICMP probes
 - 52-byte packets
 - kernel time stamping
 - ssh / Kerberos



skitter: colored by more countries

Country Code: from mask

CH

JE

ES

IT

JP

NL

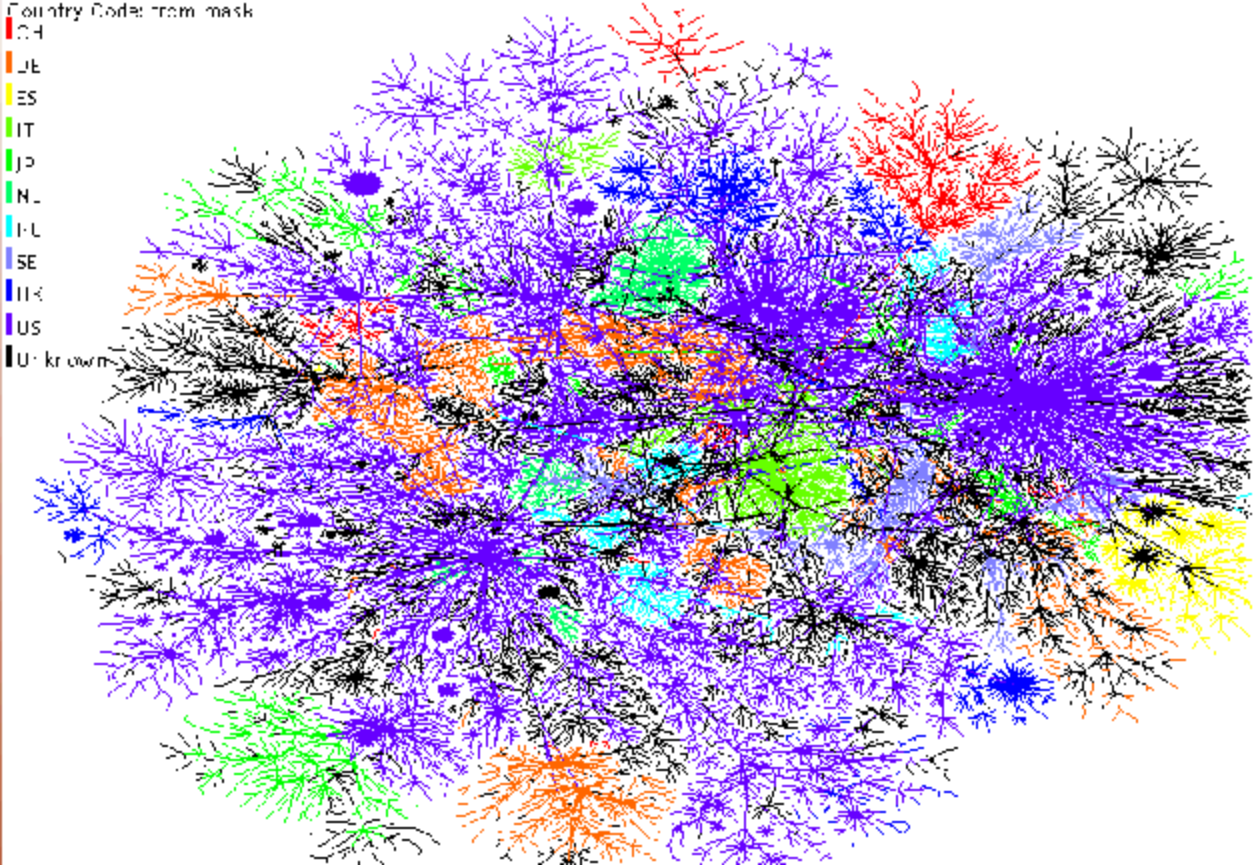
IL

SE

UK

US

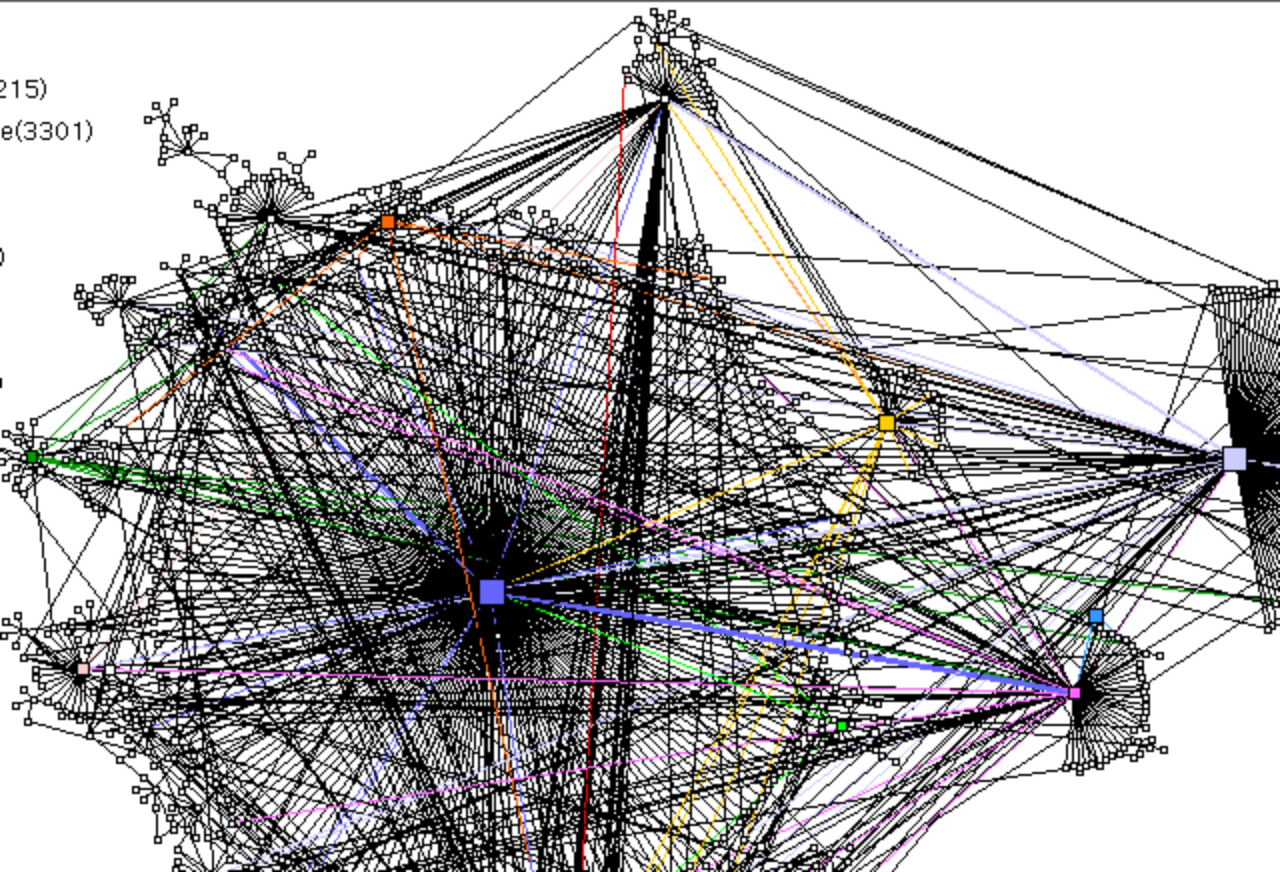
U: known



skitter: AS interconnectivity

215)

e(3301)

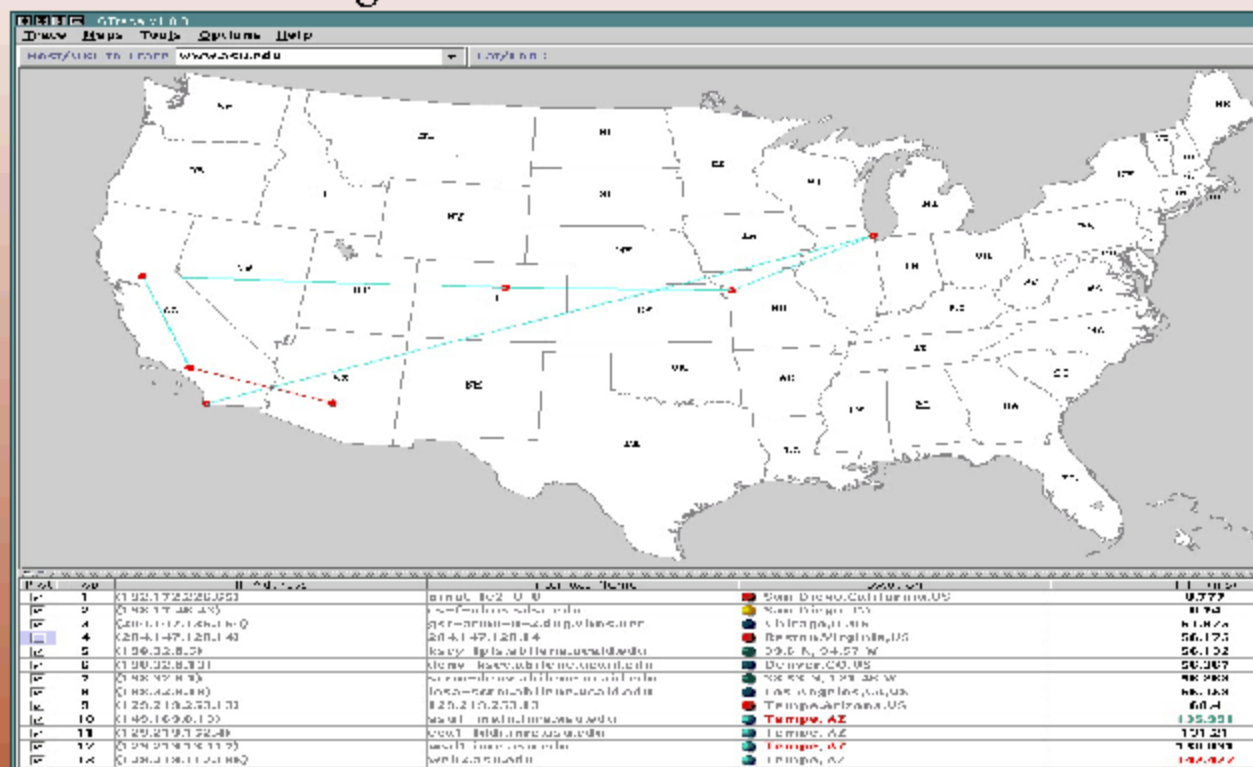


topology vis: geographic mapping

- difficult data analysis
 - requires mapping of thousands (millions?) of nodes to latitude/longitude coordinates
- NetGeo service designed to help
 - <http://netgeo.caida.org>
- backbones require company-specific heuristics
- DNS registry growth is problematic
 - no common data formats

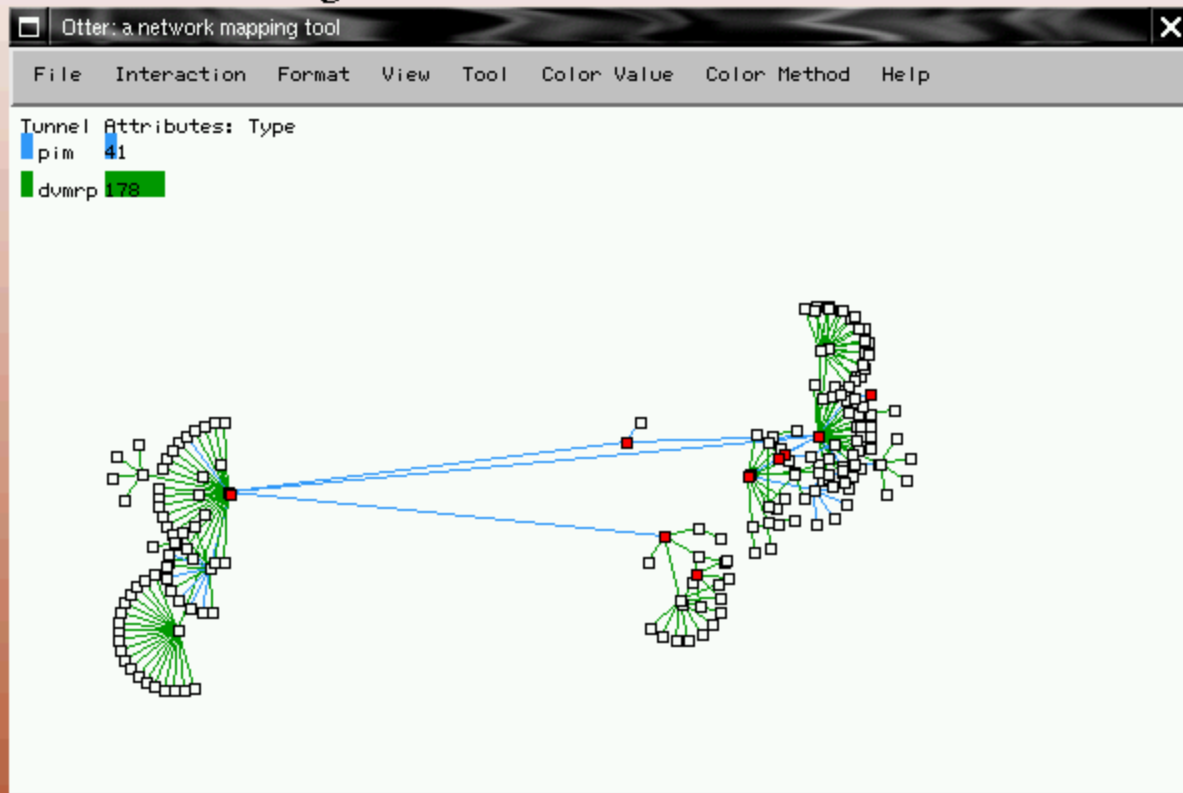
GTrace: geographic traceroute

www.caida.org/Tools/GTrace/



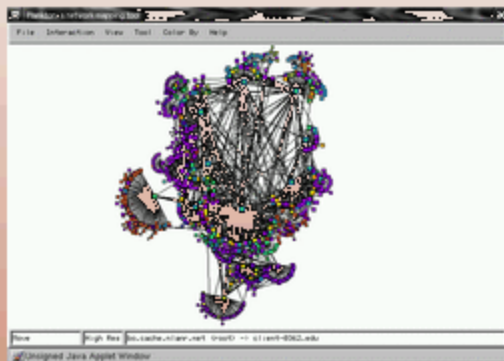
semi-geographical topology (otter)

www.caida.org/Tools/Otter



logical vs geographic topology

2-dimensional, hierarchical

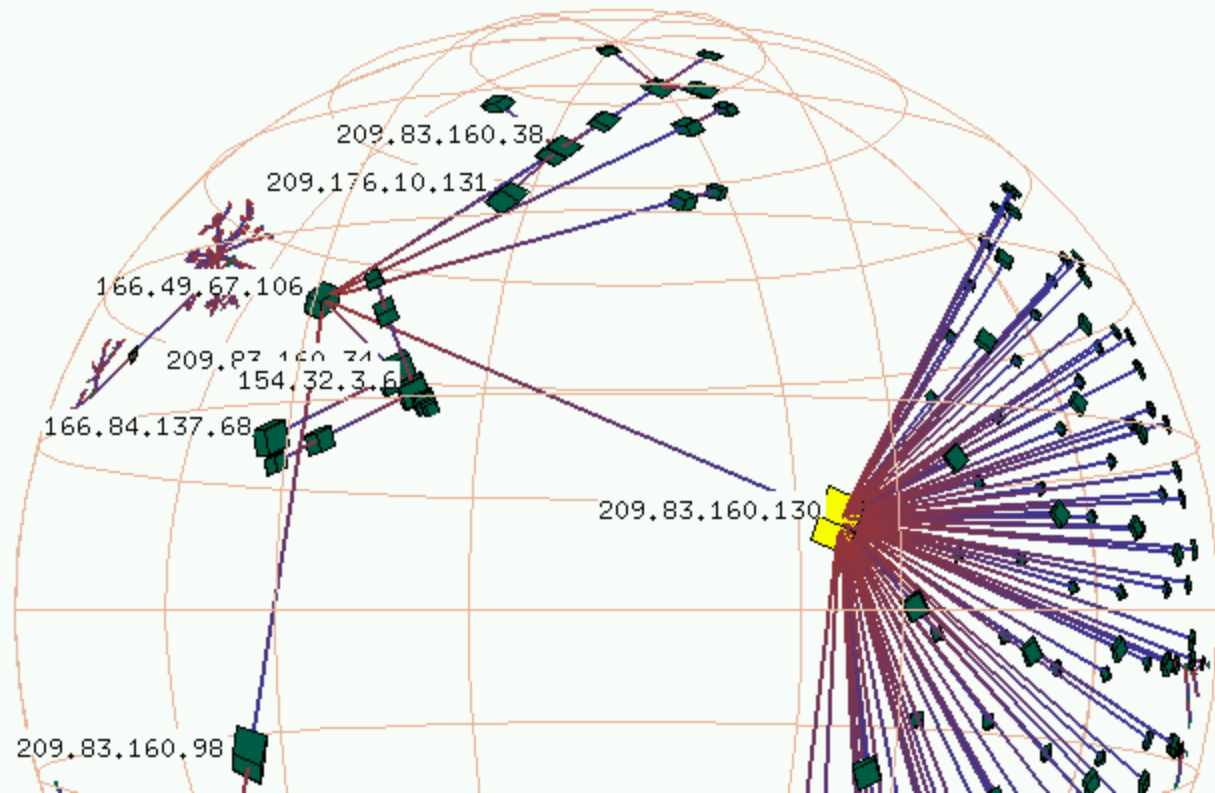


geographic



skitter: 3D hyperbolic

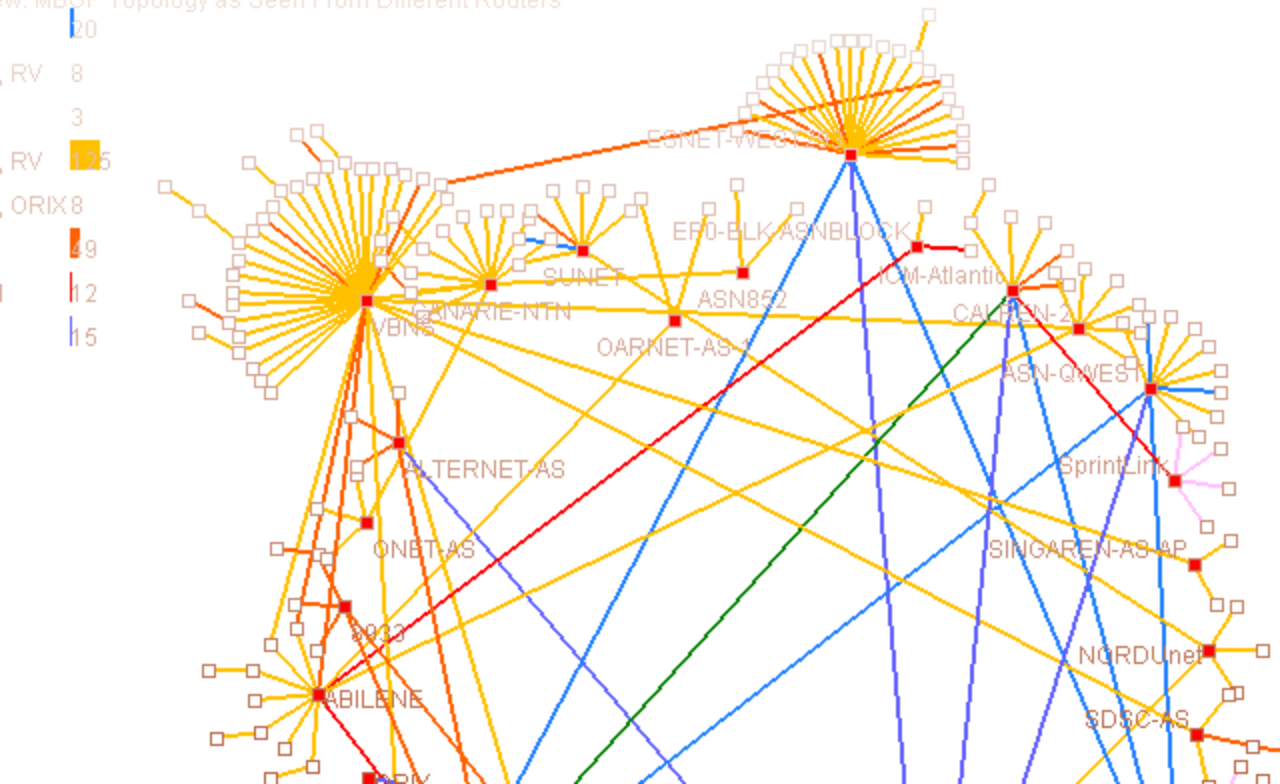
making huge topology graphs navigable



differencing routing tables

www.caida.org/Tools/Mantra (multicast)

View: MBGP Topology as Seen From Different Routers



topology vis: research priorities

- identify & extract features from large, complex datasets:
 - enable dynamic feature detection
 - develop better data aggregation/reduction techniques
 - create meaningful displays, user-friendly tools
 - accurately correlate different datasets

- obstacles:
 - mapping IP addresses to ... anything meaningful

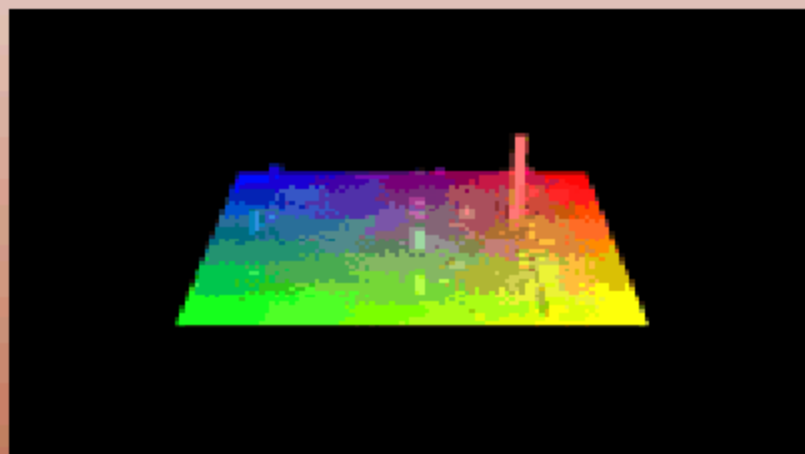
=> visualizing Internet workload

- many uses
 - workload profiling
 - performance and QOS assurance across ISPs
 - accounting/billing
- measurement tools
 - router-based (cflowd, netflow)
 - stand-alone monitors (coral,skitter)
- traditional vis can be helpful
 - time series, box&whisker, etc.
 - 3D real-time flow analysis
- evolution requires use
 - envisioning new methods?
 - better data correlation tools are essential

traffic workload by protocol

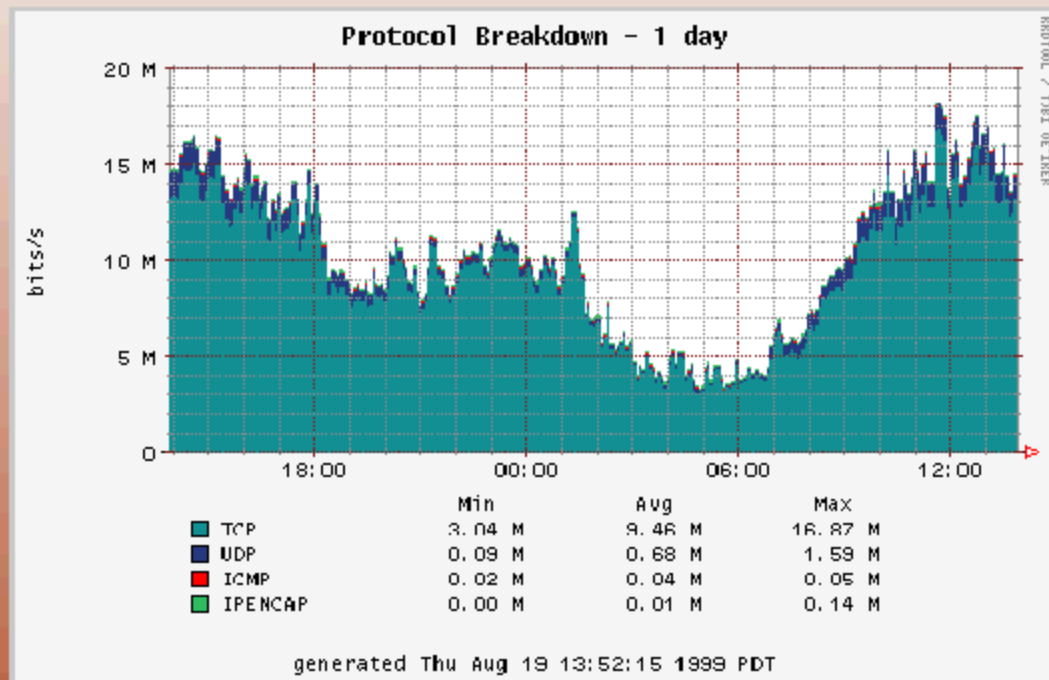
Cichlid animations

<http://moat.nlanr.net/Cichlid/>



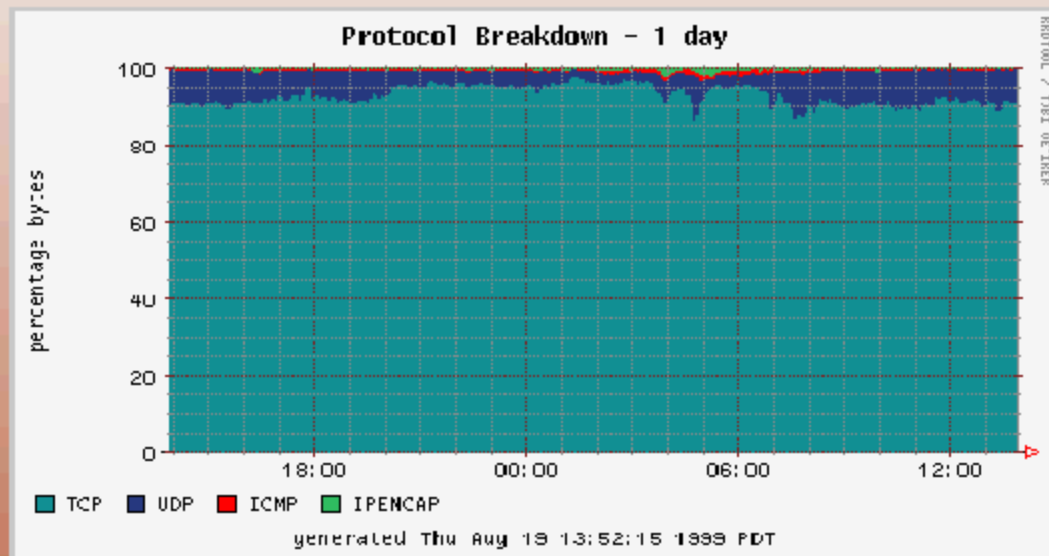
traffic workload by protocol

19 aug 99, ucsd-cerfnet



workload by protocol proportions

19 aug 99, ucsd-cerfnet

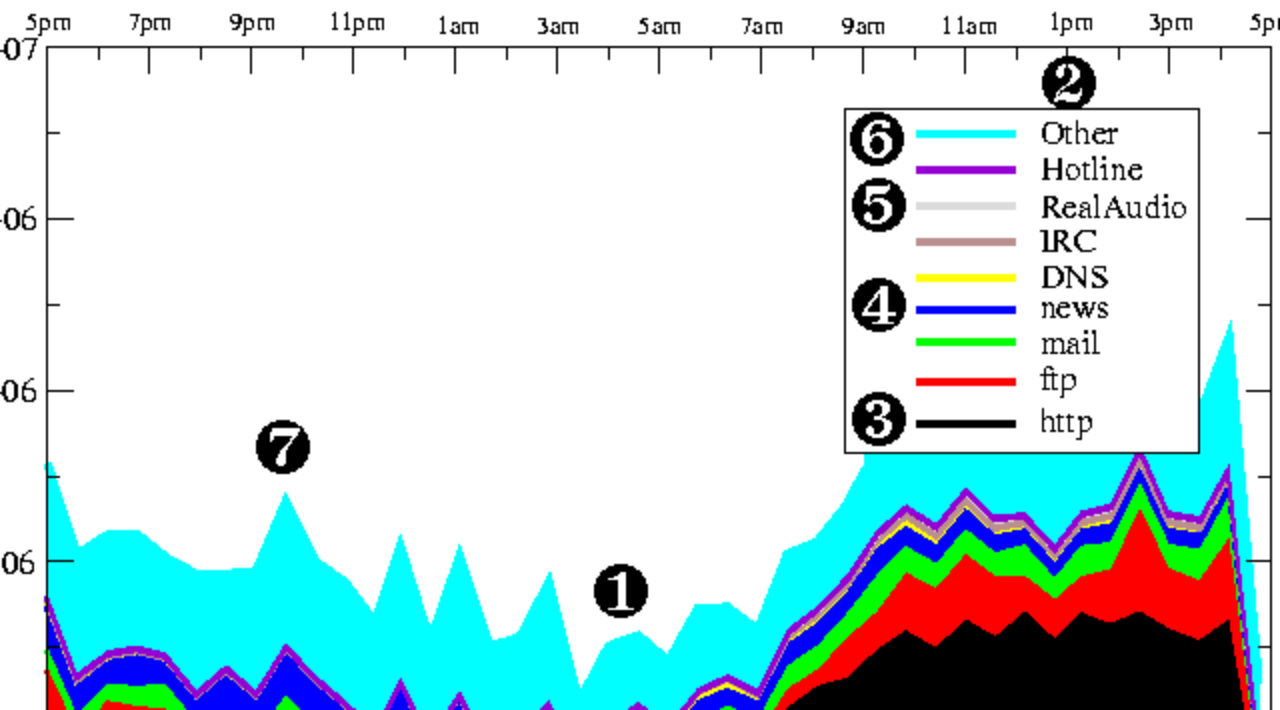


visualizing workload by applications

(e.g. ucsd-cerfnet)

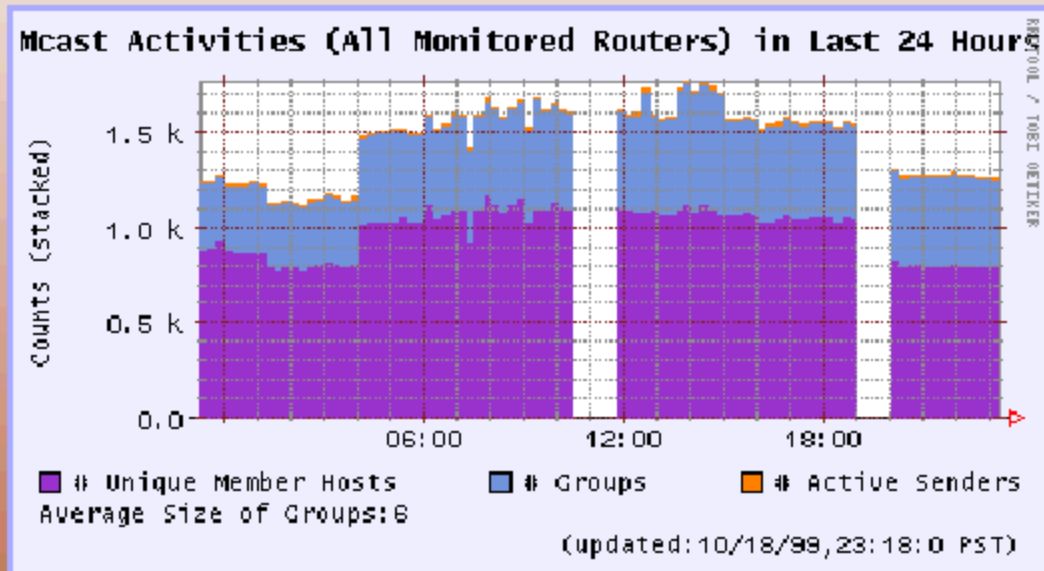
Internet applications for June 13, 1999

Local Time



multicast workloads (using mantra)

18 oct 99, fix-west.mbone.nasa.gov



workload vis: research priorities

- id and present 'useful' workload metrics, particularly given persistence of fire-fighting environment
- id significant patterns, timeframes, correlations
 - vary by user need
 - change as technologies and 'net change
- obstacles:
 - limited access to commercial networks
 - network performance impact
 - faster speeds and changing transport technologies complicate data acquisition and processing

=> performance evaluation vis

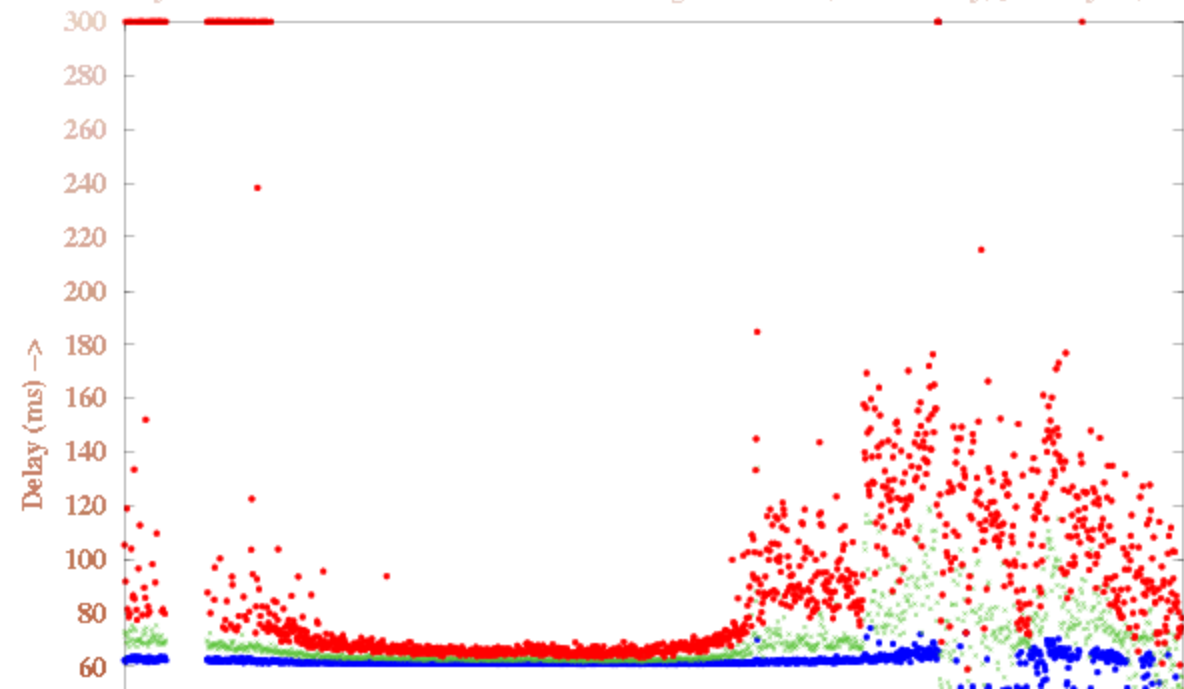
useful for:

- diagnosing problems
- verifying service level agreements
- predicting software HCI for network apps
- providing Internet weather reports

max delay
percentile delay
percentile delay

Advanced to Oregon Gigapop

Delay statistics over 1-minute intervals starting 00:00 UTC, Wednesday, January 05, 2000



perf. eval: skping (www.freebsd.org)

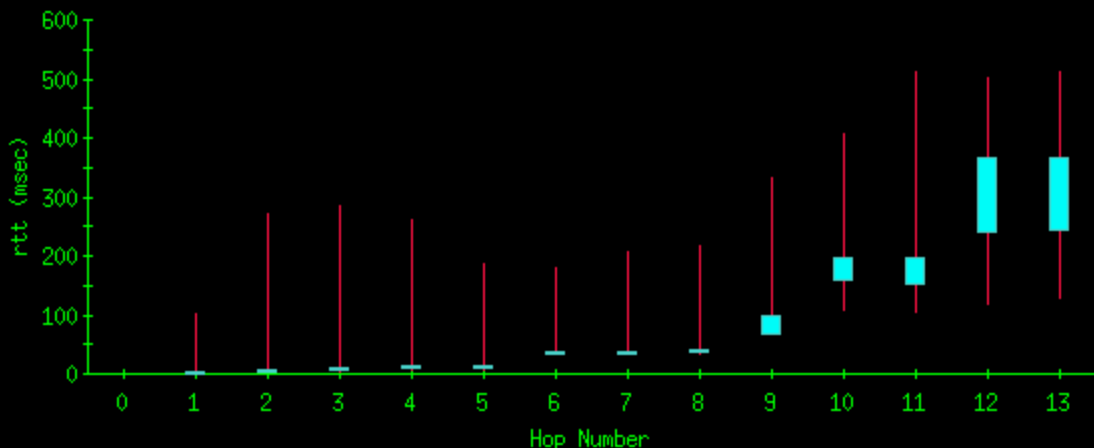
sktracegui

File

Scatter

Candle

max/95th/25th/min



Path Information

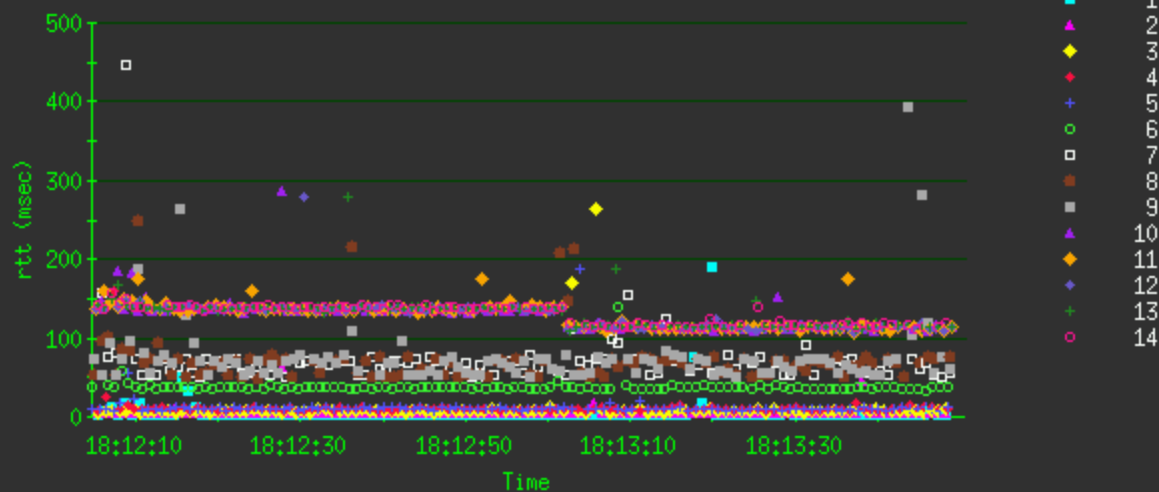
hop 1 aaqw-e0.caida.org

perf.eval: routing (path change)

sktracegui

Scatter

Candle

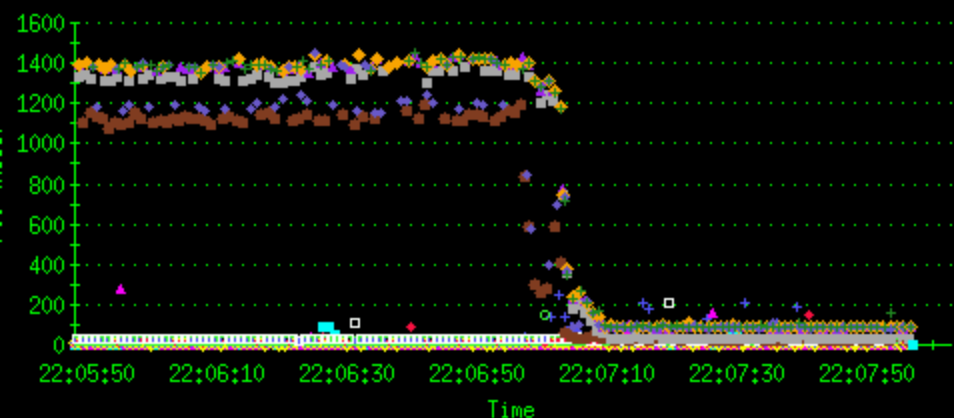


Path Information

hop 1 aagw-e0.caida.org

perf.eval: sktrace (www.cnet.com)

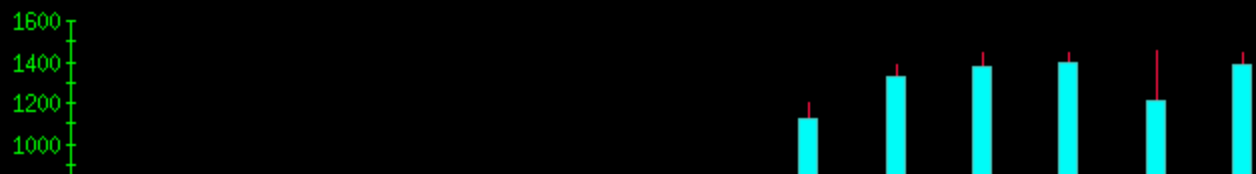
racegui



- 1 204,212,...
- ▲ 2 204,212,...
- ◆ 3 204,212,...
- ♦ 4 198,87,2...
- + 5 131,103,...
- 6 131,103,...
- 7 157,130,...
- 8 146,188,...
- 9 146,188,...
- ▲ 10 146,188,...
- ◆ 11 146,188,...
- ♦ 12 146,188,...
- + 13 137,39,1...

■ max/75th/2

(sec)



performance eval vis: priorities

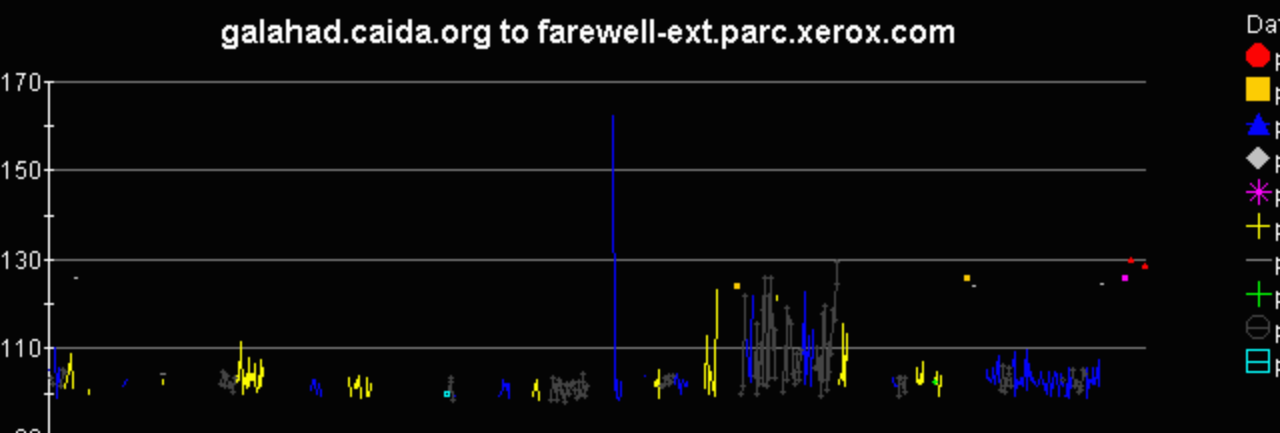
- faster data collection, faster processing, faster rendering
- need intuitive graphic presentations
 - correlating:
 - performance across sources
 - comparisons w/topology, workload, routing analyses
- obstacles
 - poorly defined user requirements/interfaces
 - negative perceptions regarding quality and worth driven by explosive growth

=> visualizing routing efficiency

- sausage/laws....
- how best to identify outages, flaps, critical paths
- how best to meet real-time challenges
- currently non-intuitive, incantation-driven

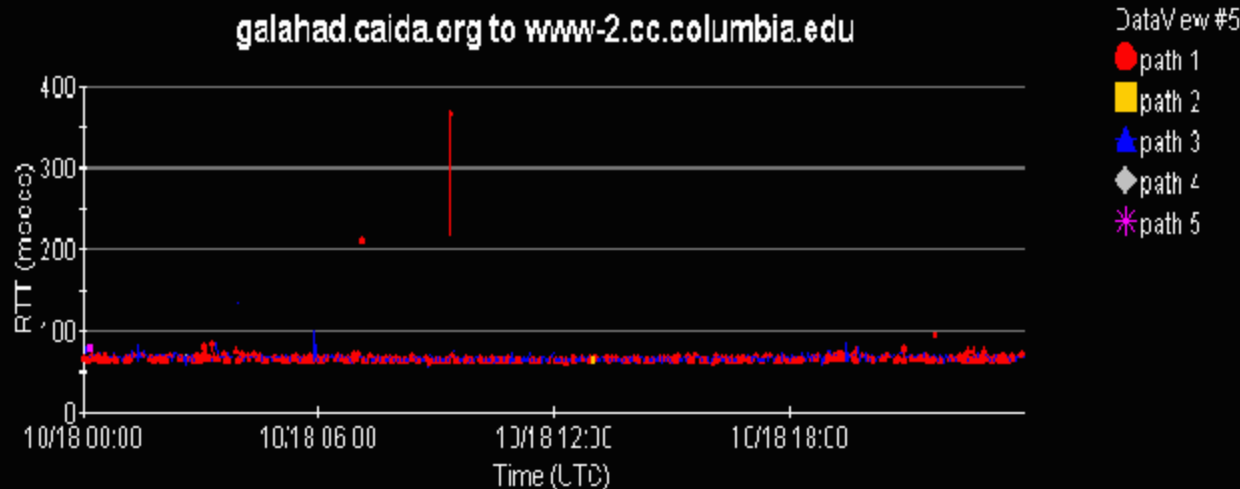
routing: example (instability)

- RTT data changes color if path changes
- 10 unique paths over 24 hour period
- lots of jitter in data
 - unlikely to be intentional
- heavy tails predominate



routing: example (load balancing)

- RTT similar over predominantly 2 paths
- likely intentional load balancing

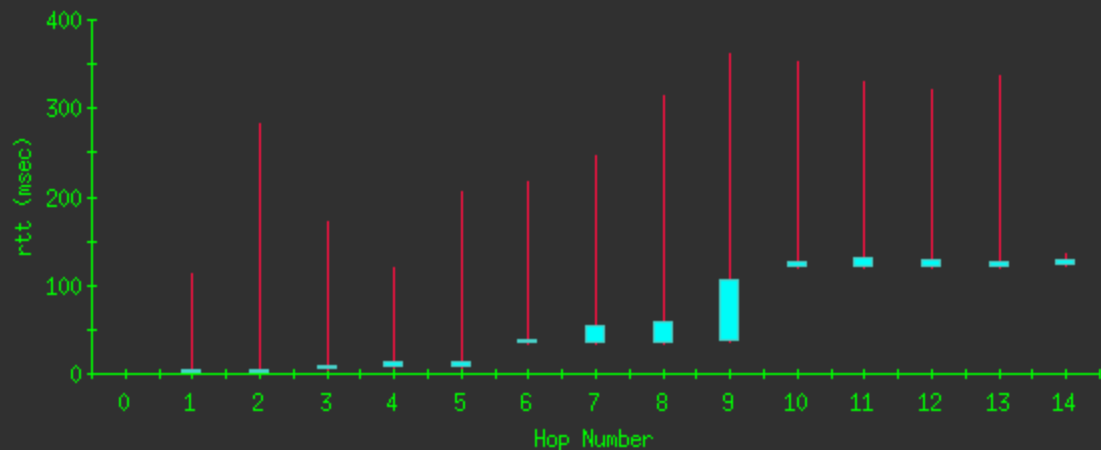


routing: sktrace (parc.xerox.com)

sktracegui

Scatter **Candle**

max/95th/25th/min



Path Information

1	aagw-e0,caida.org		
2	rtr2-ser2-0,blackrose.org		

routing vis: research priorities

- better IP routing instrumentation
- real-time analysis without interfering with performance
- realistic inter-domain routing models

- tasks
 - identification & vis of flaps, outages, critical paths
 - correlation with performance problems
 - analysis of asymmetries and routing instabilities
 - analysis of effects of unicast/multicast incongruities

routing vis: research obstacles

- routes may change faster than ability to measure or analyze
- poorly instrumented infrastructure (new tools needed)
- mapping IP address to anything (deja vu)
- prudent security dictates inhibiting research

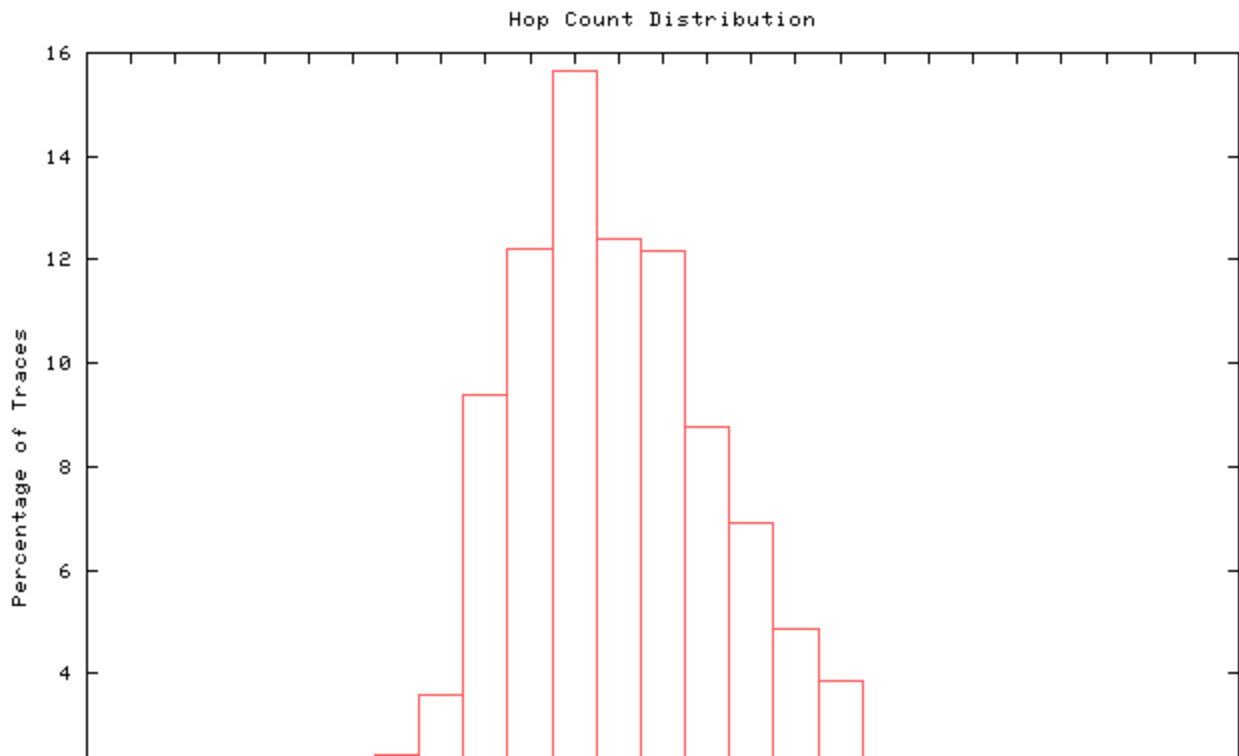
now what?

- the ideal:
 - well-instrumented infrastructure
 - seamless integration of variety of data sources
 - important for simulation/prediction
 - but unlikely for the foreseeable future

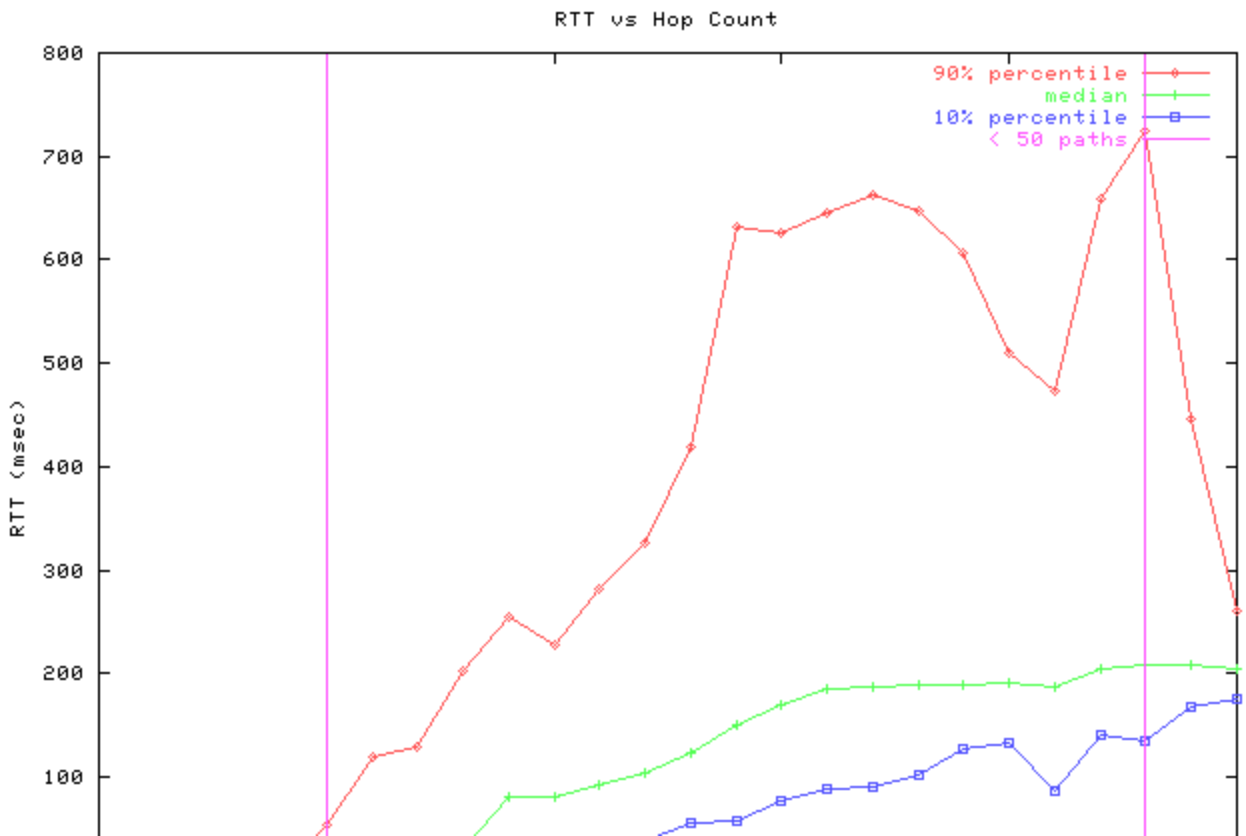
- a case study (real-life problem)
 - optimize root name server location
 - use visualization to correlate data sources / types
 - use collaborative project to encourage proactive participation
(network operators, researchers, others)

=> skitter: macroscopic case study

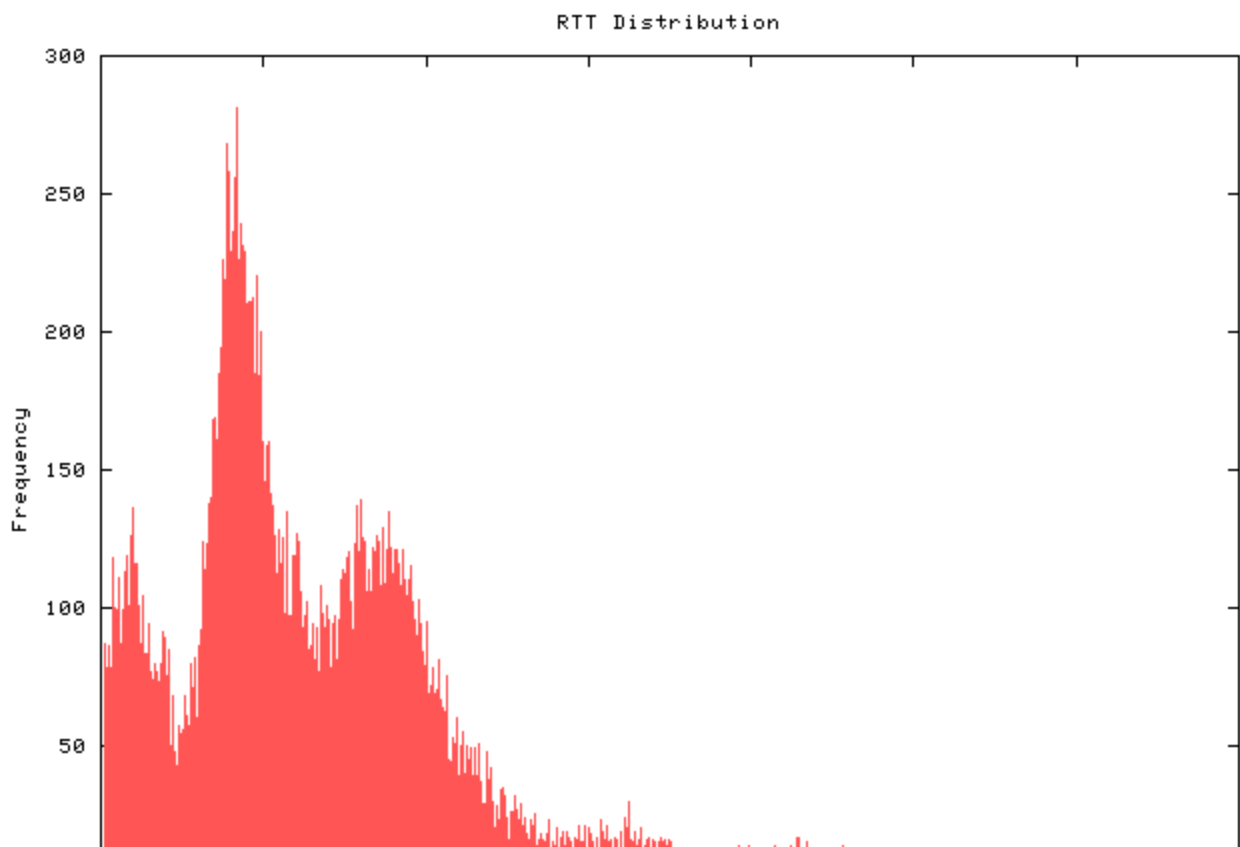
DNS f root server (pv's): path wingspans
www.caida.org/Tools/Skitter



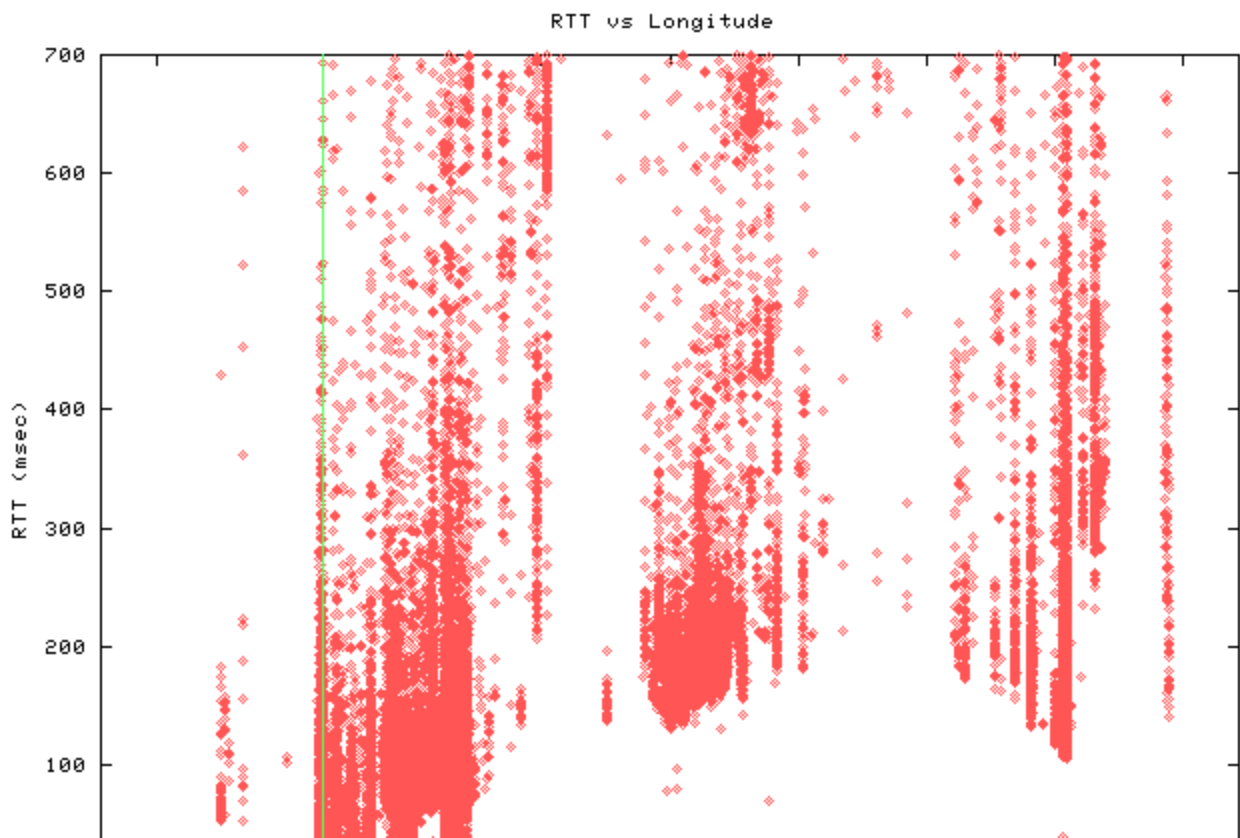
skitter: rtt vs hopcount (correlation?)



skitter: rtt distribution: tri-modal

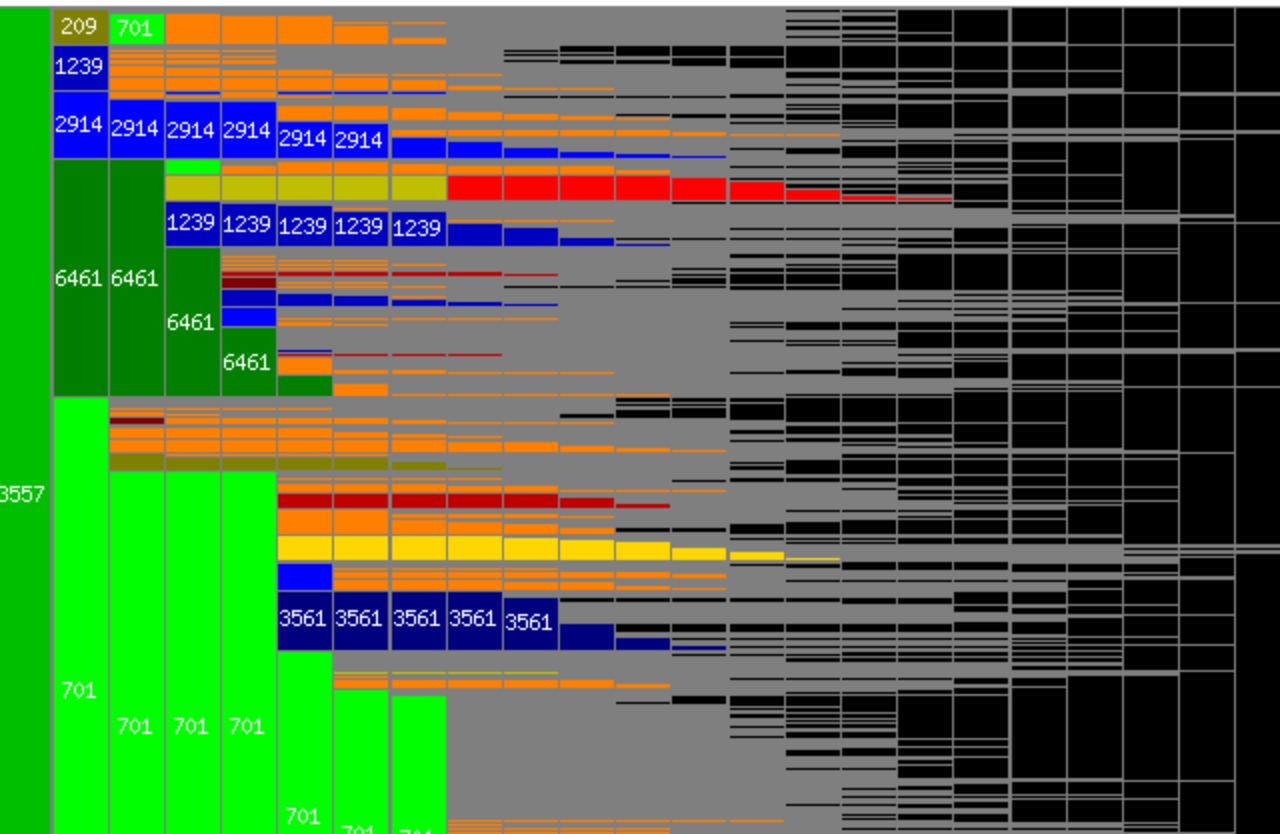


skitter: rtt vs longitude (light cone)

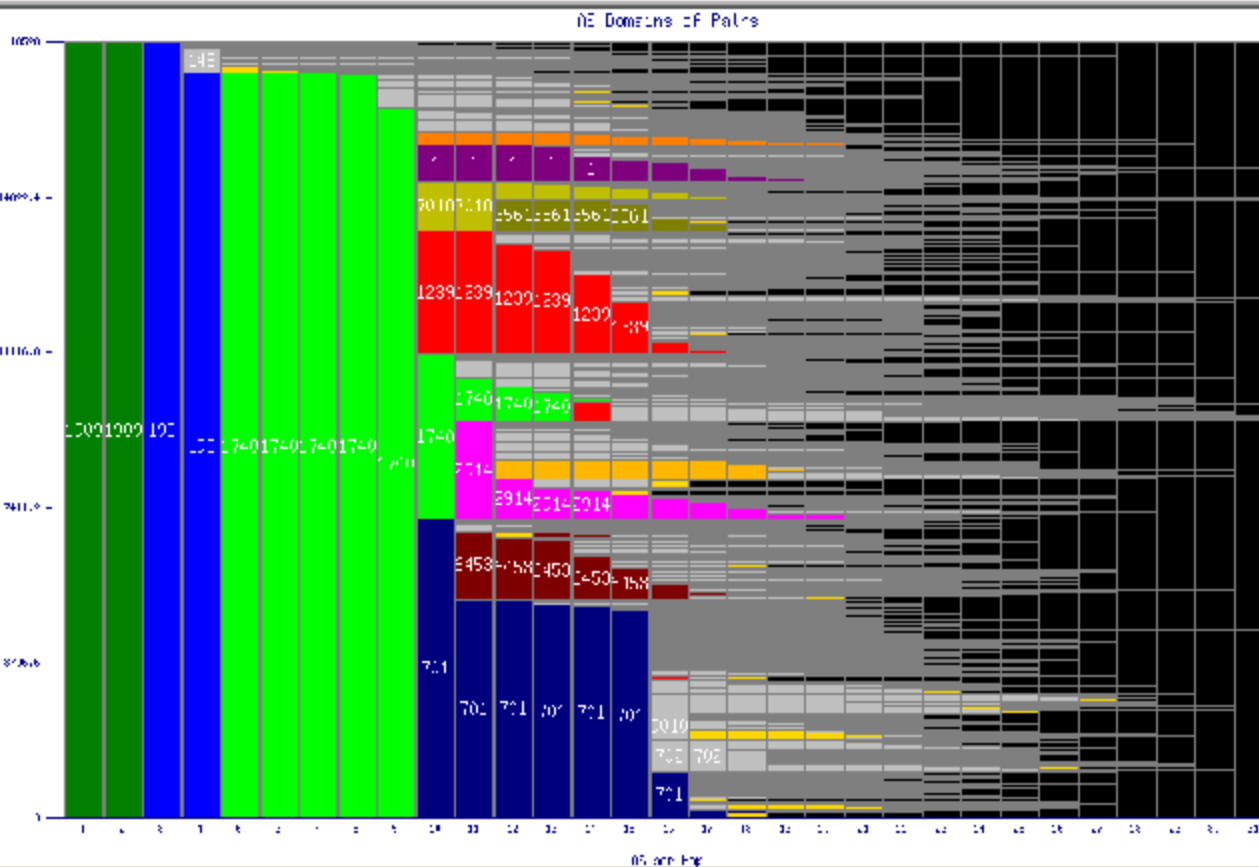


dispersion among ASes across paths

AS Domains of Paths



dispersion among ASes across paths (sdsc)



CAIDA research pushes boundaries

- analysis of complex conditions
- management of large datasets
- correlation between different datasets
- development of timely, insightful visualization

new CAIDA Vis group will investigate

- tool integration
- user interface improvements
- networked collaborative environments

potential payoffs

- give insights to **vendors** re next generation hw/sw requirements
- calibrate **user activity** (e.g., monitoring service level agreements)
- provide diagnostic and planning tools for **ISPs**
- identify and quantify infrastructure dependencies for **researchers**



caida

www.caida.org/Presentations/

kc claffy
UCSD/SDSC/CAIDA
kc@caida.org
www.caida.org