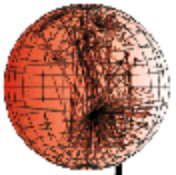


tracking a metamorphic infrastructure:

observations on our (in)ability to accurately predict, analyze or even measure conditions on the global Internet

25 may 2000
sane 2000 maastricht.nl

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

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

abysmal but unsurprising

- little capacity to predict, depict, or even measure traffic behavior on current and advanced networks
- few tools to engineer/operate networks or identify traffic anomalies in real time
- doesn't stop researchers from building junk
- doesn't stop random users from doing random junk (no dearth of activity)
- increasing risk to infrastructural potential

Internet measurement taxonomy

- topology (circulatory/respiratory)
- performance (physiology/psychology)
- workload (cardiovascular/GI)
- routing (neuroscience)

correlation
.....(holistic Internet measurement?)

topology: existing initiatives

macroscopic, IP layer

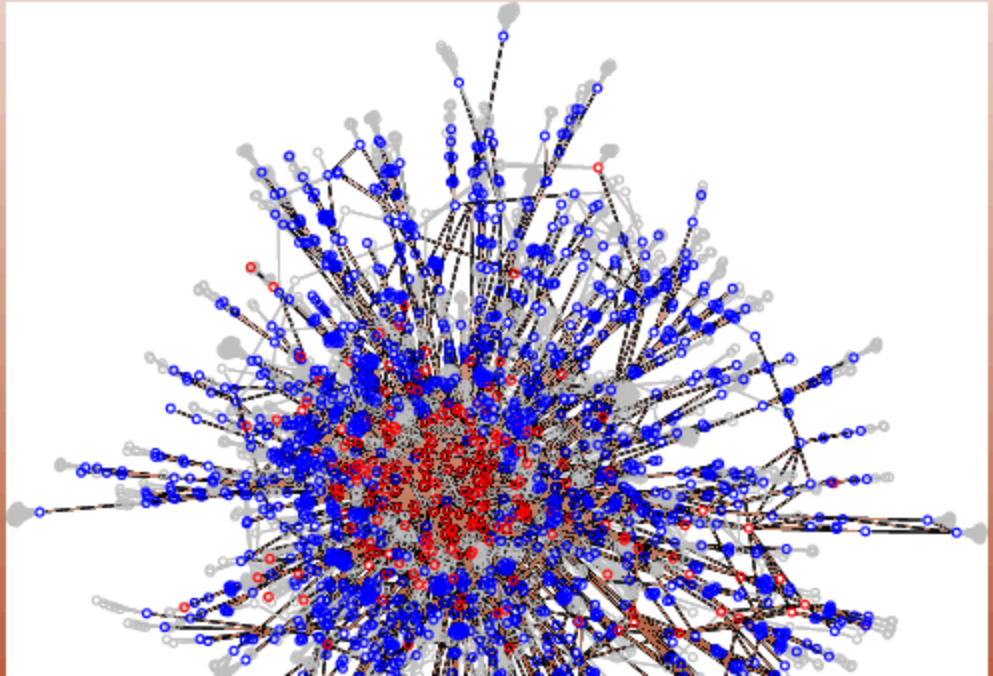
- USC/ISI: mercator
- Lucent: burch/cheswick maps
- CAIDA: skitter

why do we care?

- traffic engineering
- track global growth/change
  - arguable: increased potential/manageability
- correlation with
  - routing
  - performance
  - workload

topology: usc/isi's mercator

- informed, random address probing
- single host, some src routing
- 3 wks, 7/99, 150k interfaces, 200k links



topology: Burch/Cheswick maps

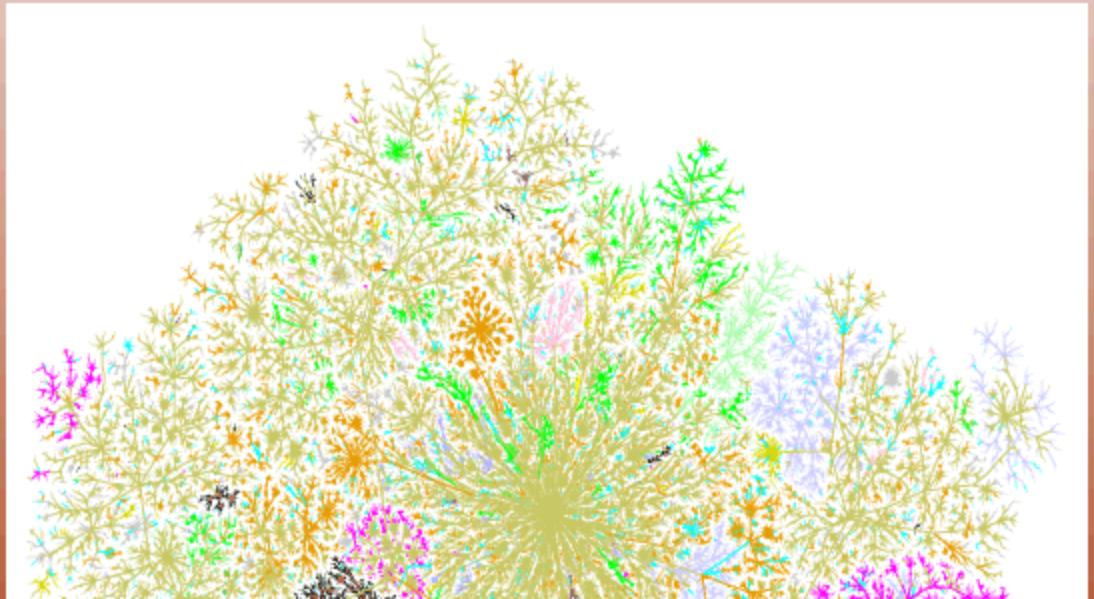
- UDP-traceroute from single host
- 1 IP addr per prefix in core route table
- convert good IPs to hostname (new & very old, 5k a day)
- 10k networks per day, full scan 1/month
- desists to targets that complain
- daily and historical data snapshots available online

http://www.cs.bell-labs.com/who/ches/map/

topology: Burch/Cheswick maps

visualizations

- minimum distance spanning tree
- 2–D, topological
- maps by: ISP, IPaddr, TLD, hopcount
- buy big posters @peacockmaps.com

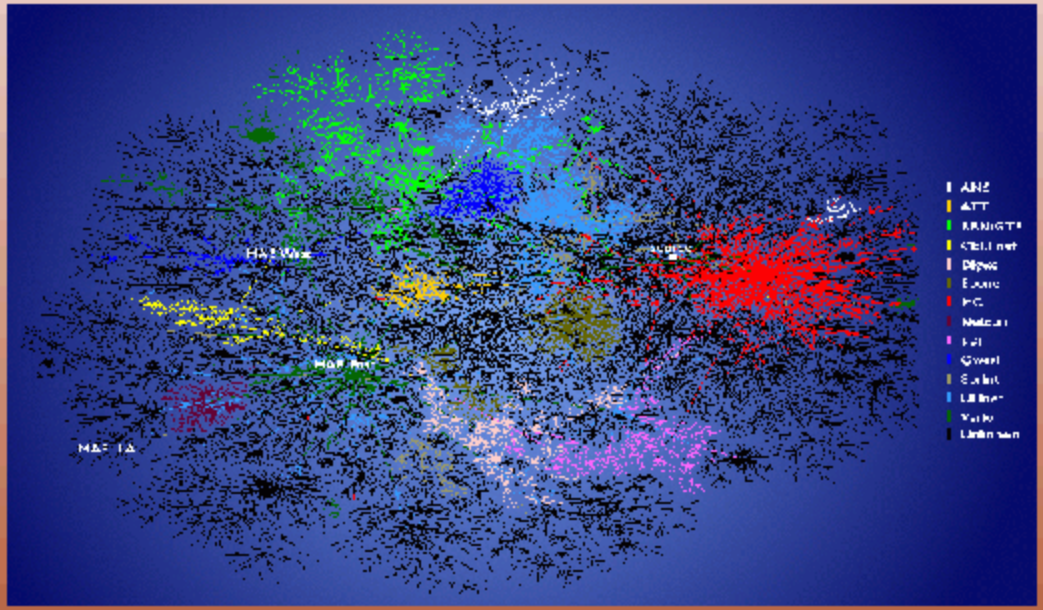


topology: caida's skitter

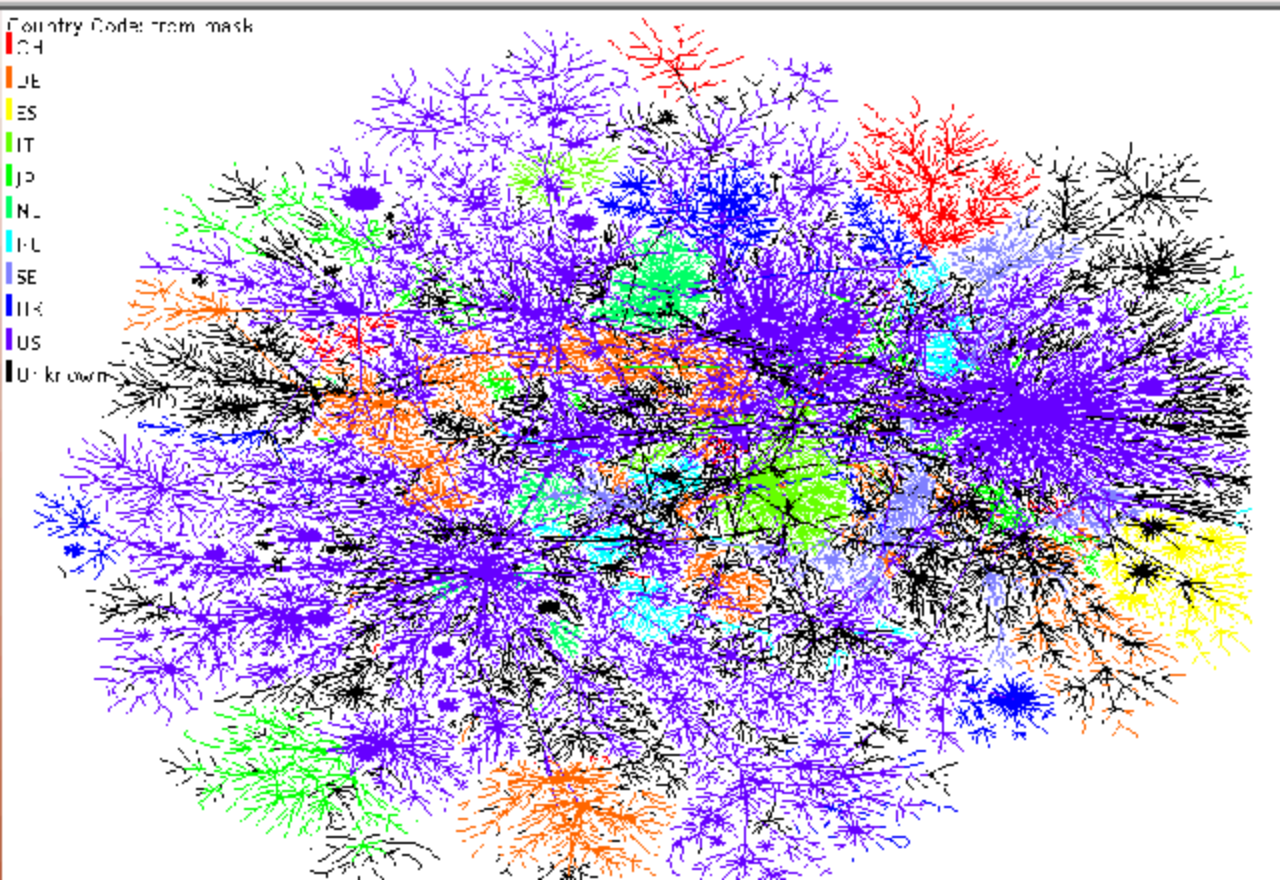
- track/depict topology cross-sections
  - 17 monitors (inc. some root name servers)
  - forward IP path and round-trip delay
  - multiple dst lists (22k servers, 36k dns)
  - remove targets that complain
- architecture
  - continuous, parallel ICMP probes
  - depending on dst list size, O(1) probes/hr/dst
  - 52-byte packets
  - kernel time stamping
  - ssh / Kerberos
- correlate path perf. w events, e.g. BGP
- identify critical pieces of infrastructure
- case studies of relevant cross-sections

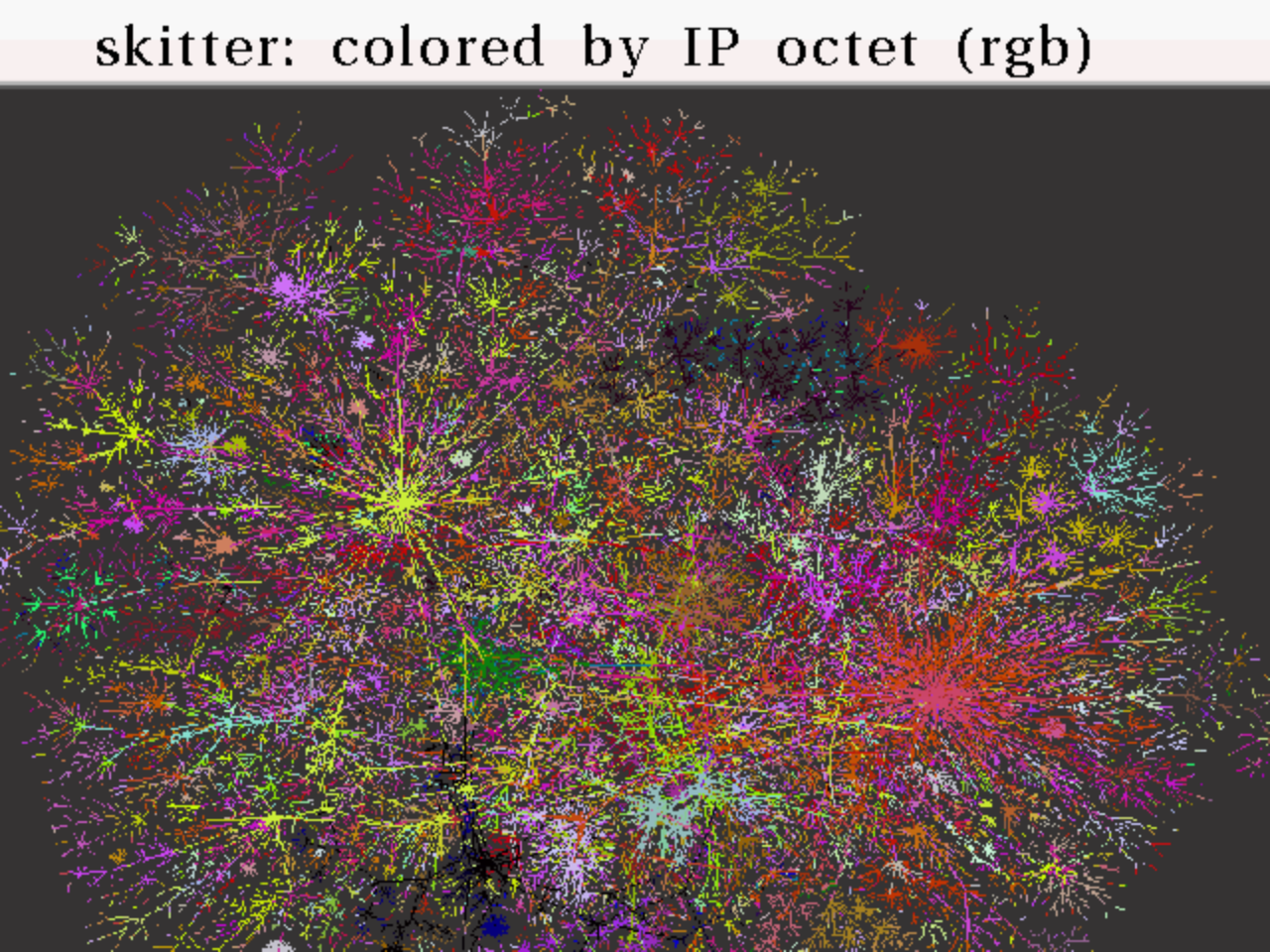
topology: skitter

- 150k nodes, 270k links in 5 days (11/99)
- lose 1% of addresses per month (30/day)
  - not repopulating main list yet
- can use hal/ches' code for 2D layout



skitter: colored by countries



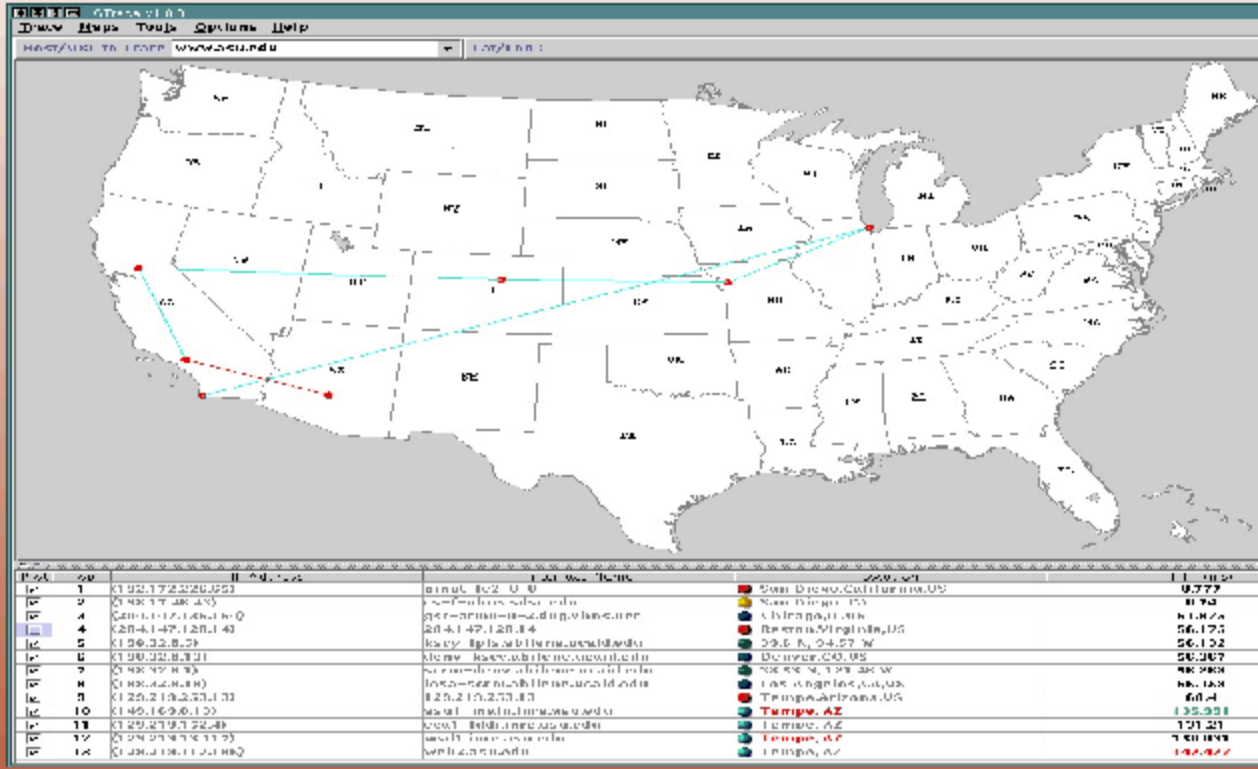

skitter: colored by IP octet (rgb)

topology vis: geographic mapping

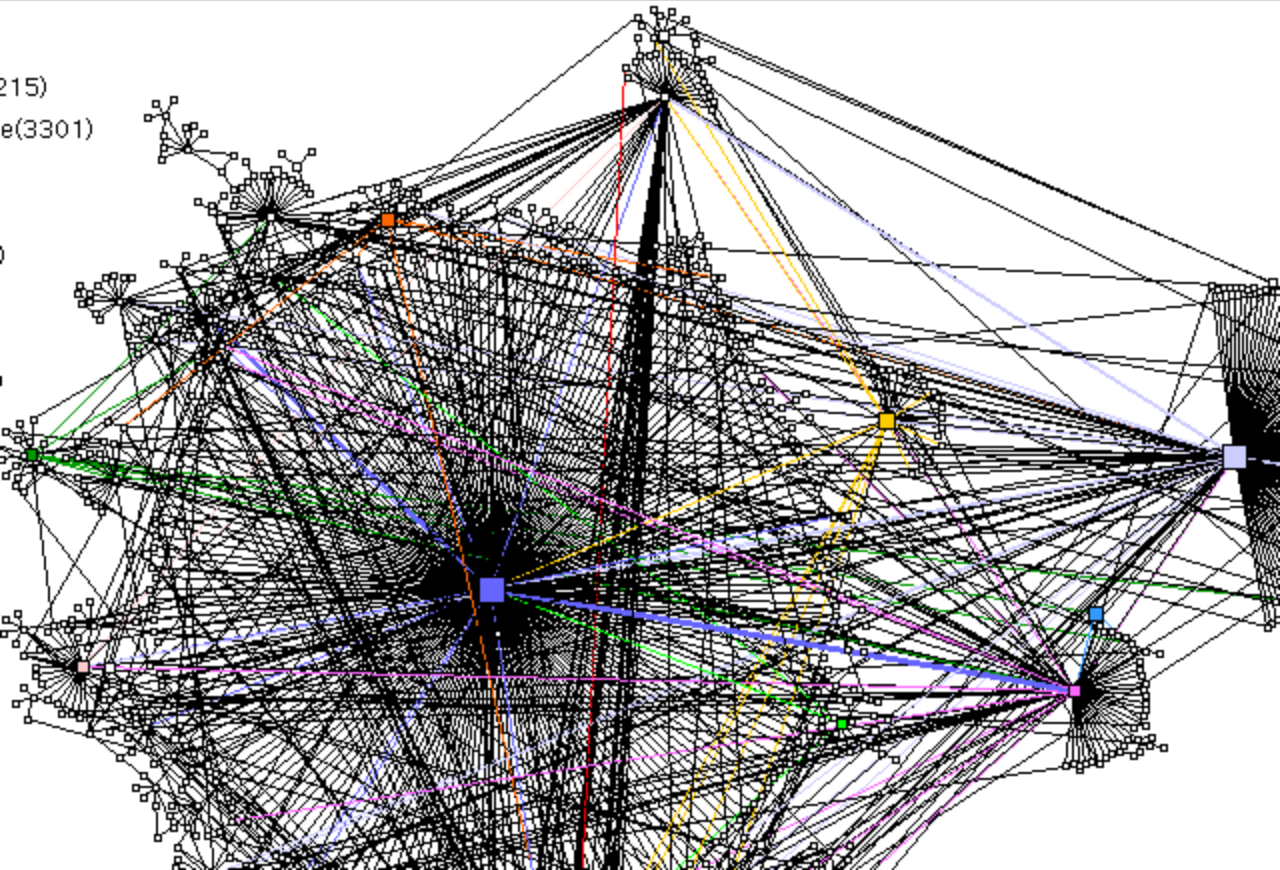
- difficult data analysis
  - requires mapping of thousands (millions?) of nodes to latitude/longitude coordinates
- NetGeo service designed to help
  - http://netgeo.caida.org
- backbones require company–specific heuristics
- DNS registry growth is problematic
  - no common data formats

GTrace: geographic traceroute

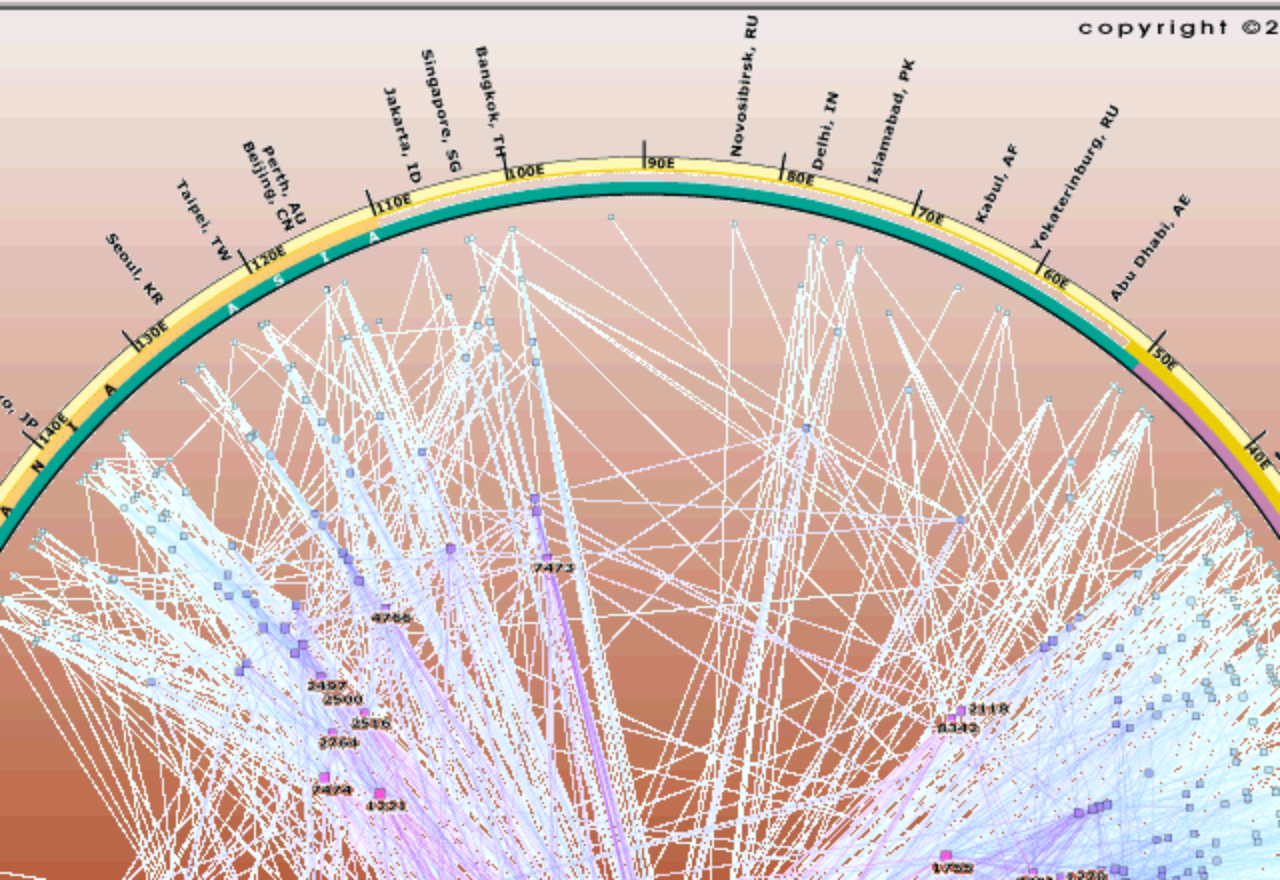
www.caida.org/Tools/GTrace/



skitter: AS interconnectivity

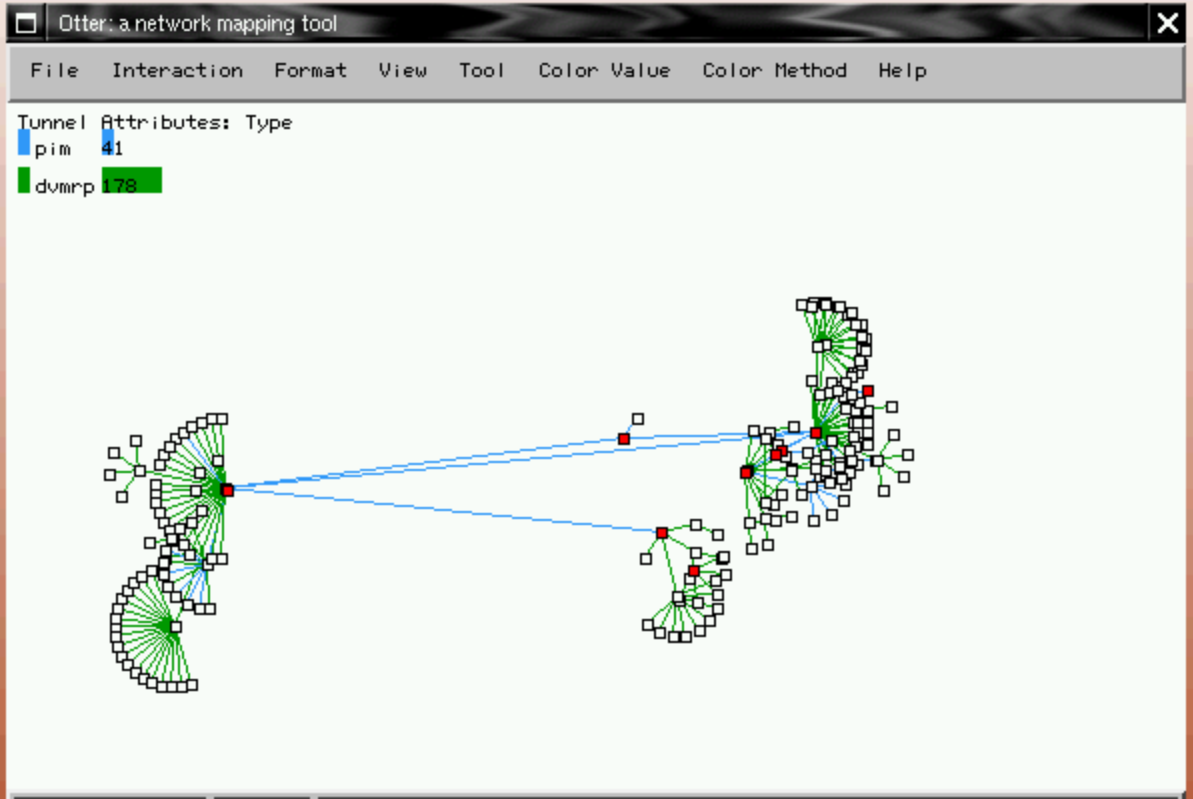


skitter: AS interconnectivity



semi–geographical topology (otter)

www.caida.org/Tools/Otter



topology/perf.: 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 (not just geography)

  things getting worse not better

topo./performance: priorities

- faster data collection, faster processing, faster rendering
- large scale public database of performance data
  - across many sources
  - comparisons w/topology, workload, routing analyses
- obstacles
  - poorly defined user requirements/interfaces
  - negative perceptions regarding quality and worth driven by explosive growth
  - uniform methodology impossible

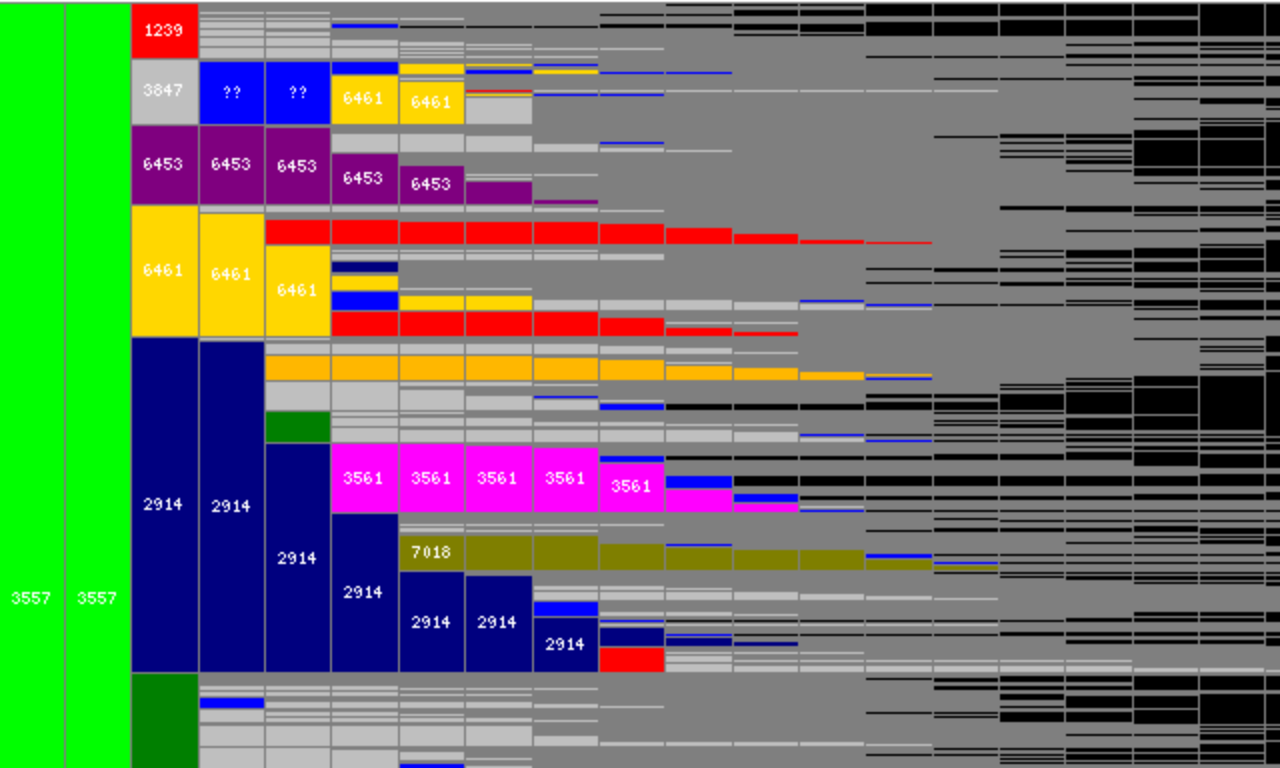
skitter case study: asia–pacific

'Measurements of Internet topology
in the Asia–Pacific Region'
http://www.caida.org/papers/

- 9 sources, about 2k dsts in asia pac
- topology, performance
- country peering, third party transit
- stability: IP vs AS level
- 60–70% of IP paths from 7 monitors are stable during day
  - closer to 40% for KR, NZ boxes
- 90% of AS paths stable
- little correlation between RTT & distance
- even less between RTT & hop count

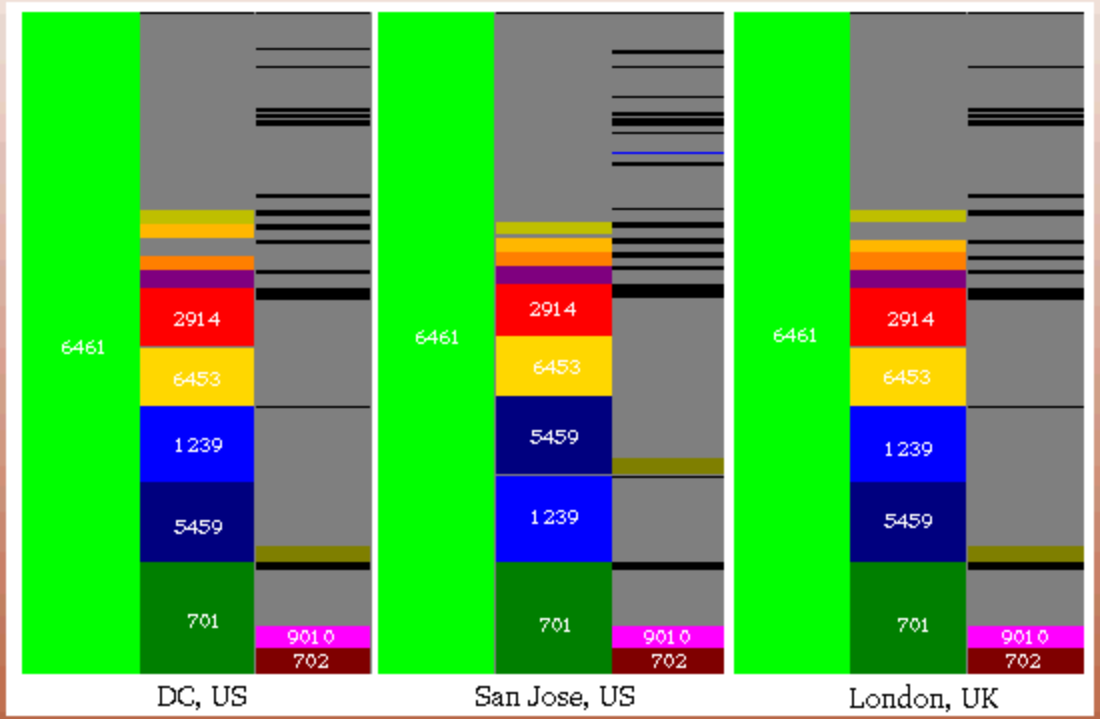
skitter: AS path dispersion

sample from palo alto source
stacked bars proportional to routes thru given AS



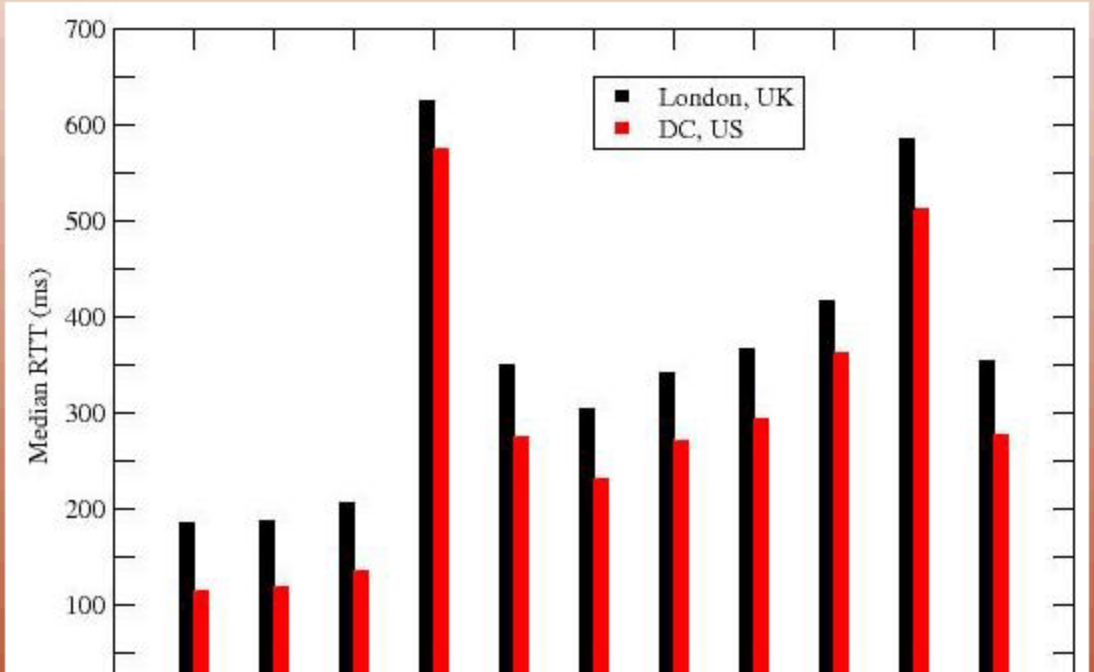
skitter case study: BGP policy for an ISP

globally consistent routes
claim to customers, verifed w skitter



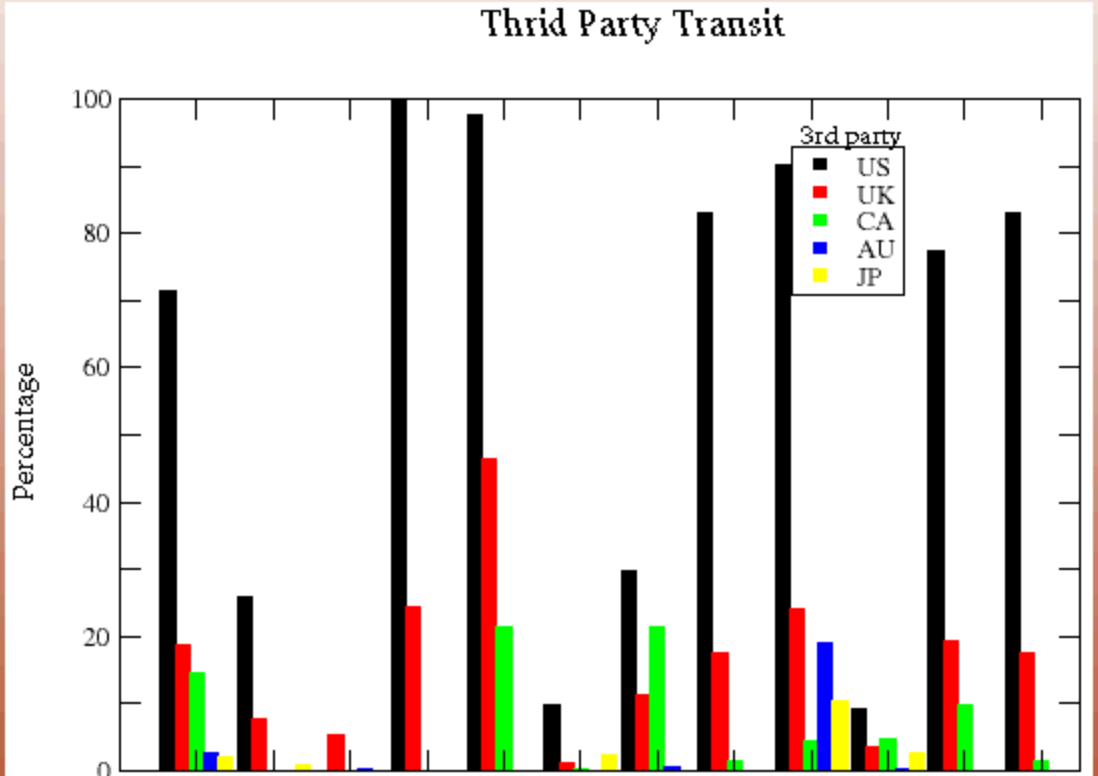
skitter case study: BGP policy for an ISP

globally consistent routes
doesn't include performance or IP routes
- most london routes probably go thru DC first
- RTT measurements substantiate



BGP policy: US as transit

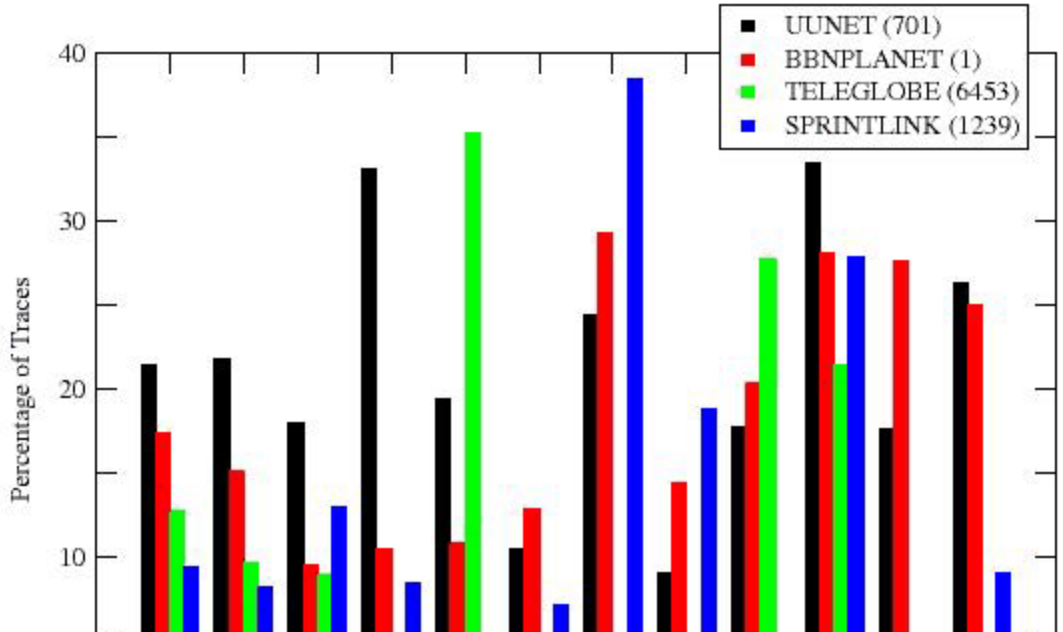
transit: neither src nor dst
can only answer for connectivity, not traffic



BGP policy: US as transit

top ASes providing transit per international dsts

- more than one can transit per path
- 4 major ISPs appear in 52% of traces (1 in .ca)
- no per-country ISP monopoly (again, traffic may look different)



BGP policy: US as transit

US transit for 71.5% of paths (not traffic!)

- 100% to MX, 98% to peru, chile
- 80–90% for cn, hk, tw, au, nz
- for most paths, US only 3rd party

2nd biggest transit country: canada

AU provided transit for 46% of all paths to NZ

0 means < 1% (blank means 0)

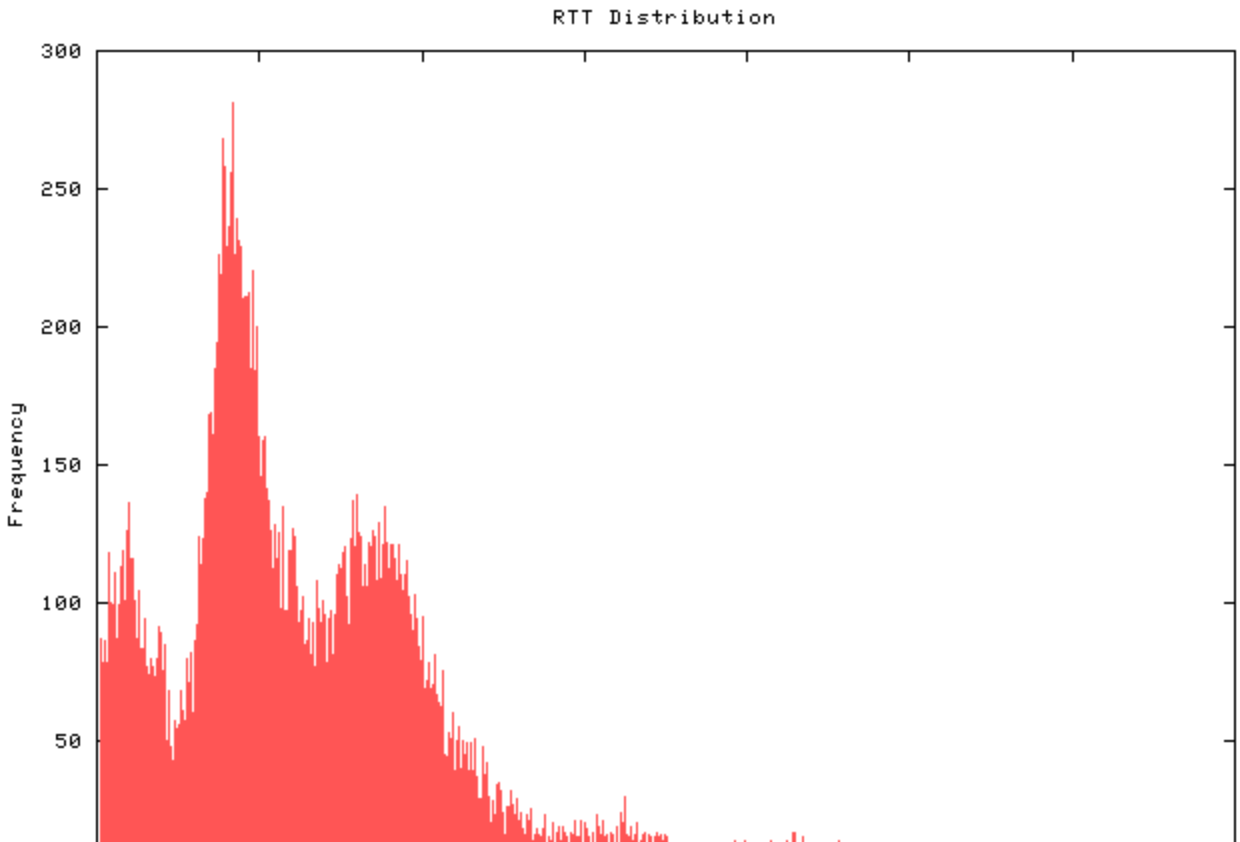
	all	AU	CA	C_H	JP	KR	MX	NZ	SEA	SWA	TW	US
US	71.5	77.8	82.0	90	49.5	61.6	100	79.6	63.0	97.8	83.5	
CA	13.3	8.3			4.9	37.5	2.1			27.5	22.3	1.3	0.2
AU	2.8				18.4					46.1	1.6		
0.4													
JP	1.2			1.4	7.4		10.5			12.0			0.3
NZ	0.9	3.7											
EUR	0.7				2.1		1.7			4.2	27.0		
UK	0.7	0.0	0.0			0.1			0.0	5.8	21.1		0.2
SEA	0.3	0.7			5.6								
AR	0.1										5.2		
AE	0.1									1.9			
CH	0.1										2.8		
MM	0.1									1.6			

skitter case study: DNS roots

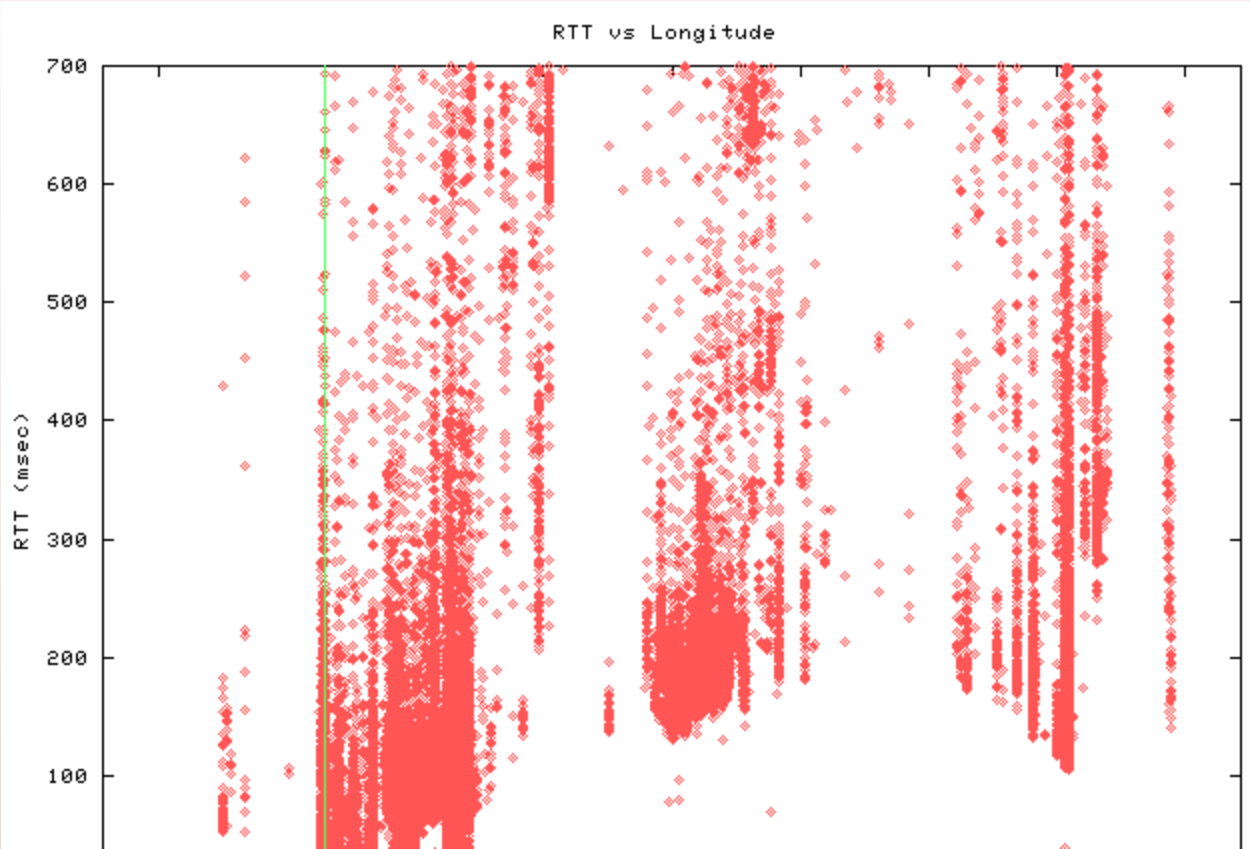
- RSSAC, DNS technical advisory committee to ICANN
- goal: optimize root nameserver location
  - co-locate skitter hosts w root servers
  - demonstrate root server performance in serving target community
  - develop techniques for evaluating architectual optimality for root server placement
  - visualization to correlate data sources/types
- use collaborative project to encourage proactive participation (network operators, researchers, others)

(www.caida.org/Tools/Skitter/RSSAC/)

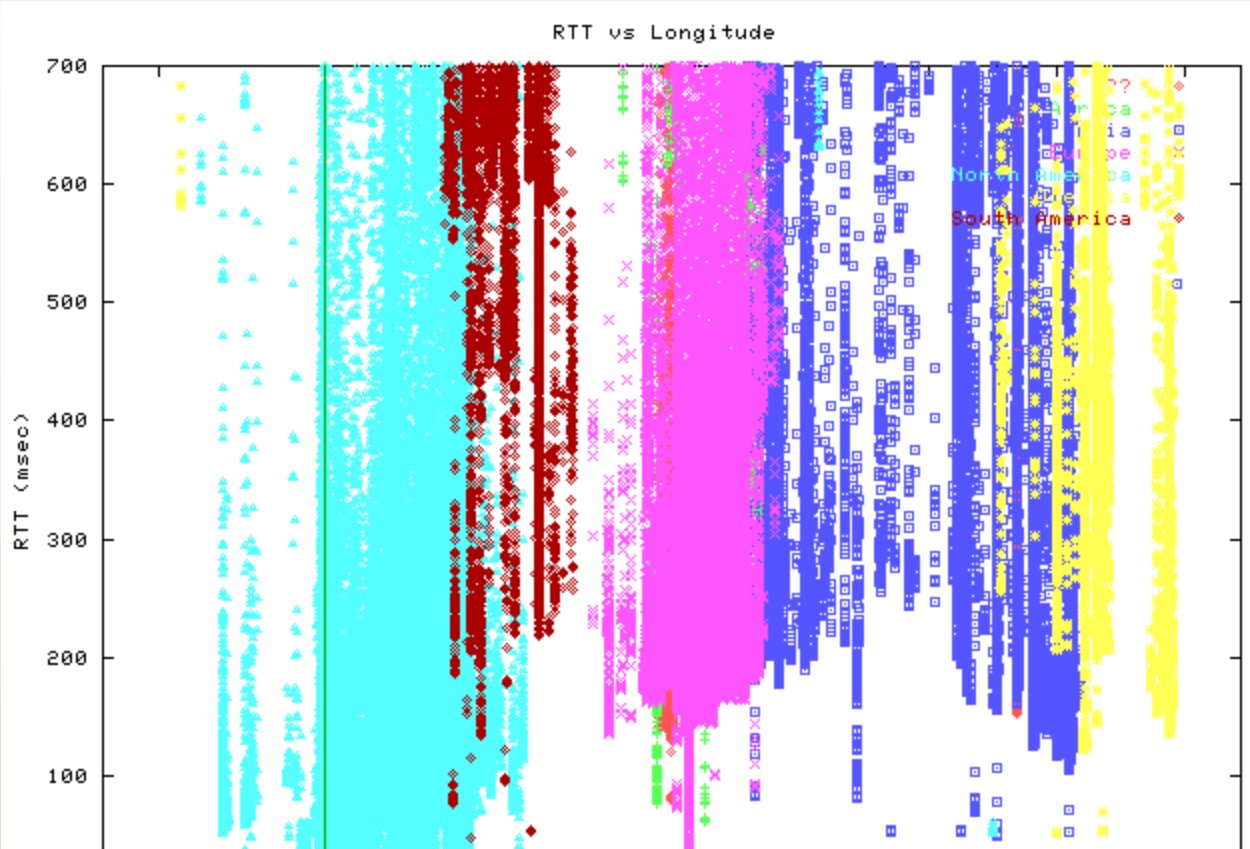
skitter: rtt distribution: tri-modal



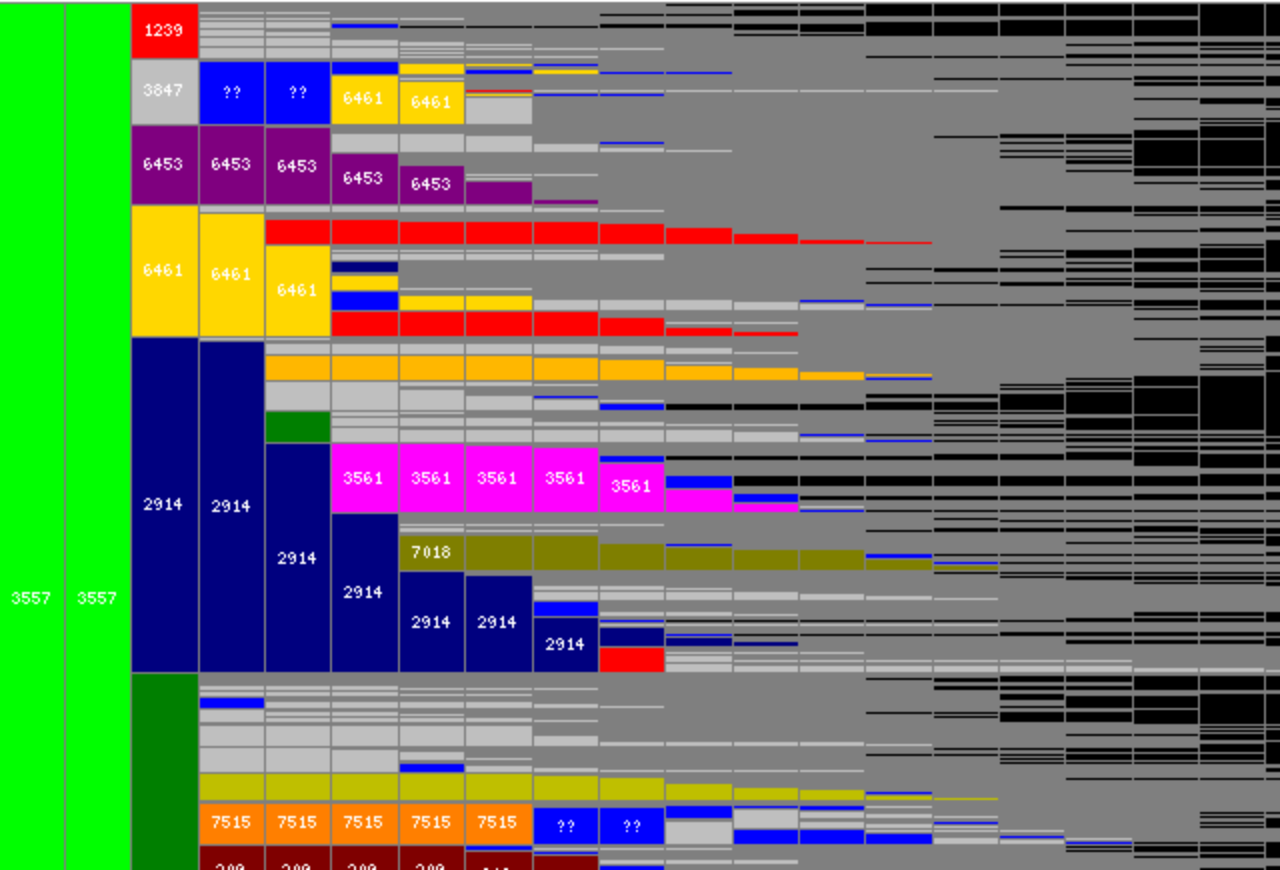
skitter: rtt vs longitude (light cone)



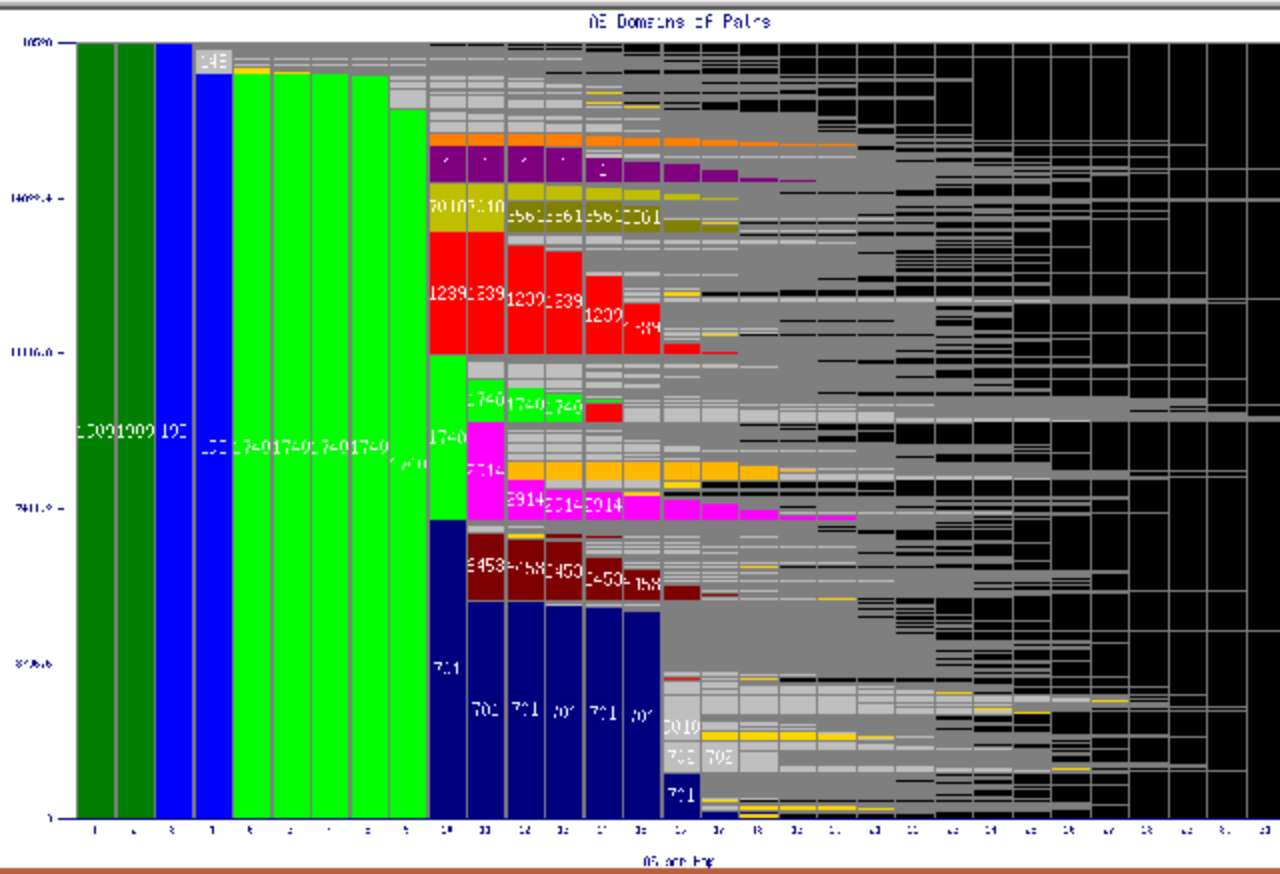
skitter: rtt vs longitude (light cone)



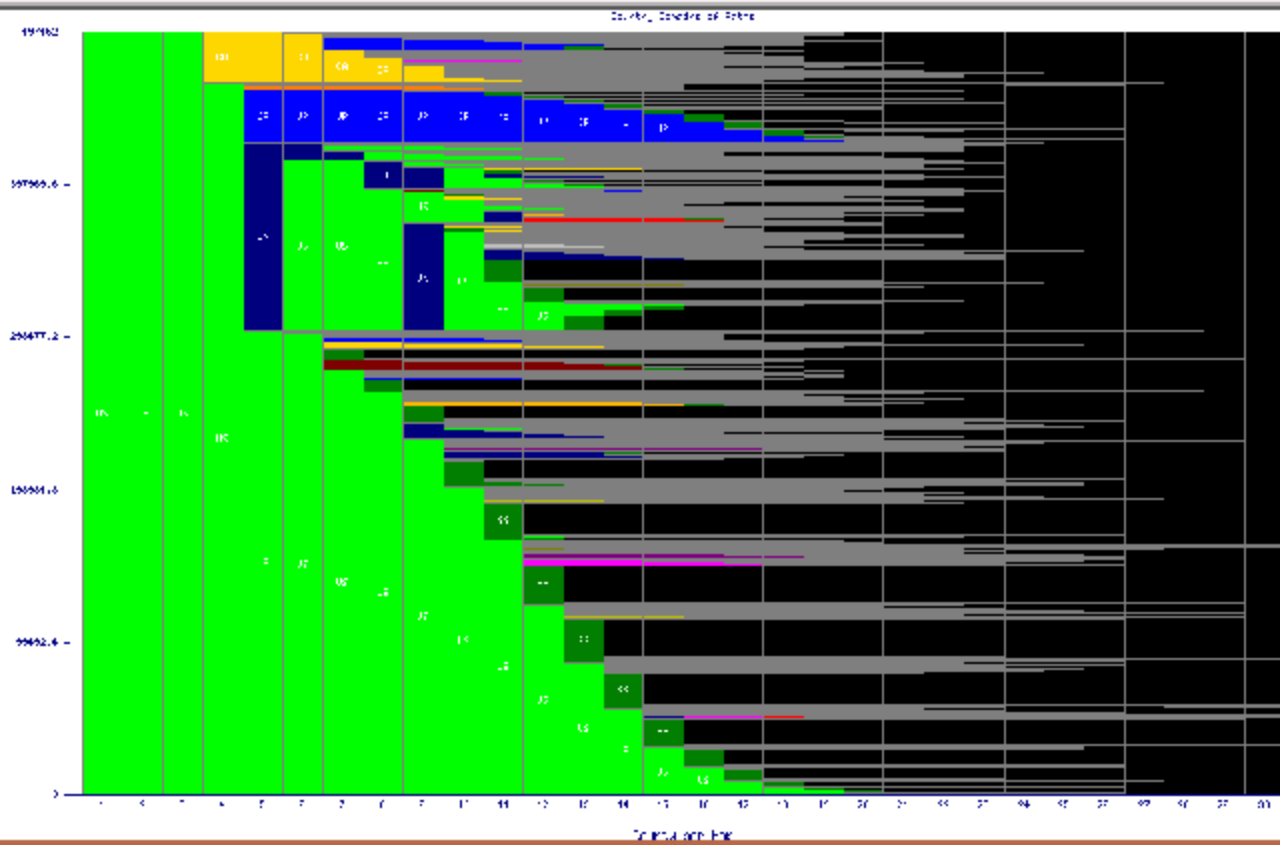
skitter: dispersion among ASes across paths



skitter: AS dispersion across paths (sdsc)



skitter: country dispersion across paths



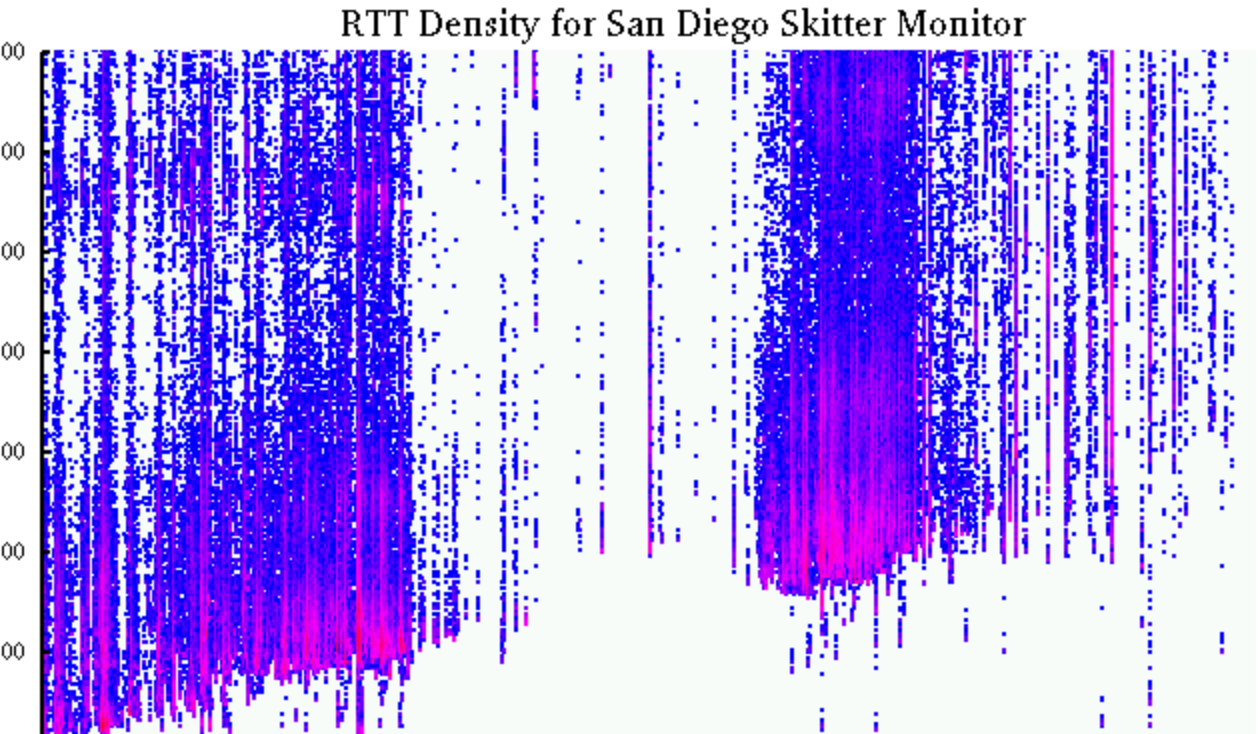
DNS roots study: future

- get other roots instrumented
- gather/analyze client lists
- correlation among different sources
- determination of connectivity metrics
  - closeness
  - redundancy
  - persistence of paths
- how many clients not secondaries
- skitter to client sets from non-root sources

skitter: other interesting possible studies

RTT versus distance

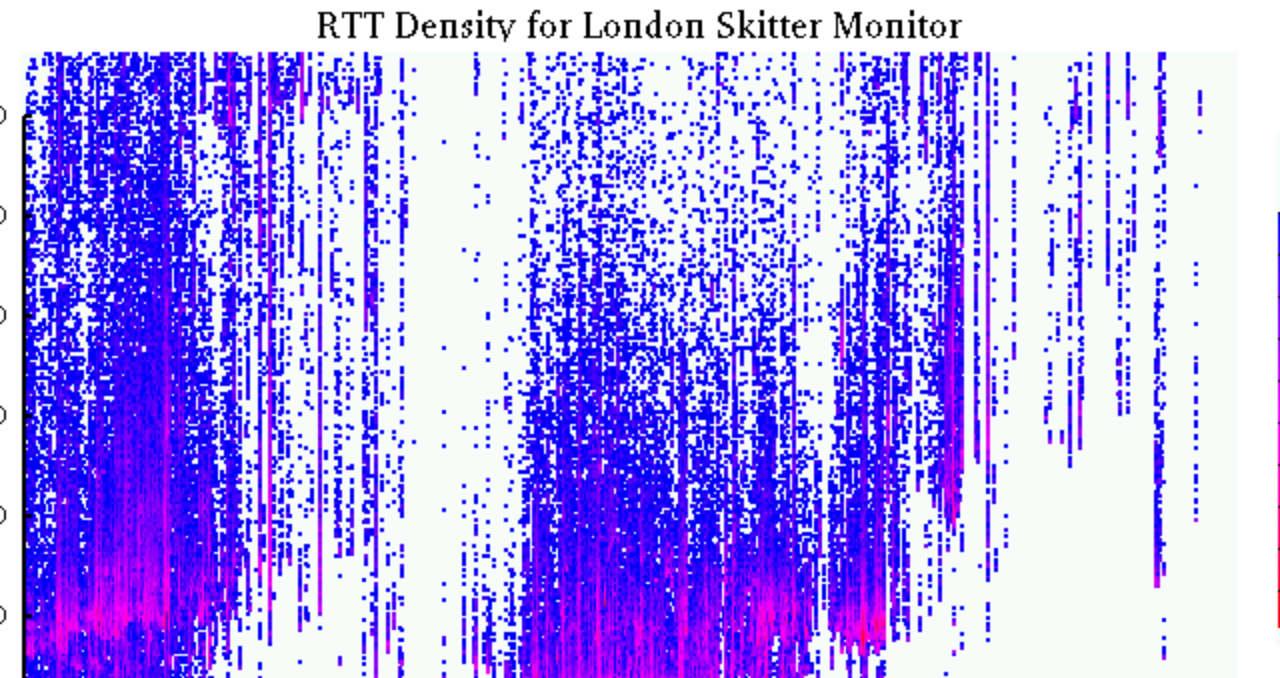
- earth circumf., X, X+Y, to-US-fr



skitter: rtt versus distance

london source

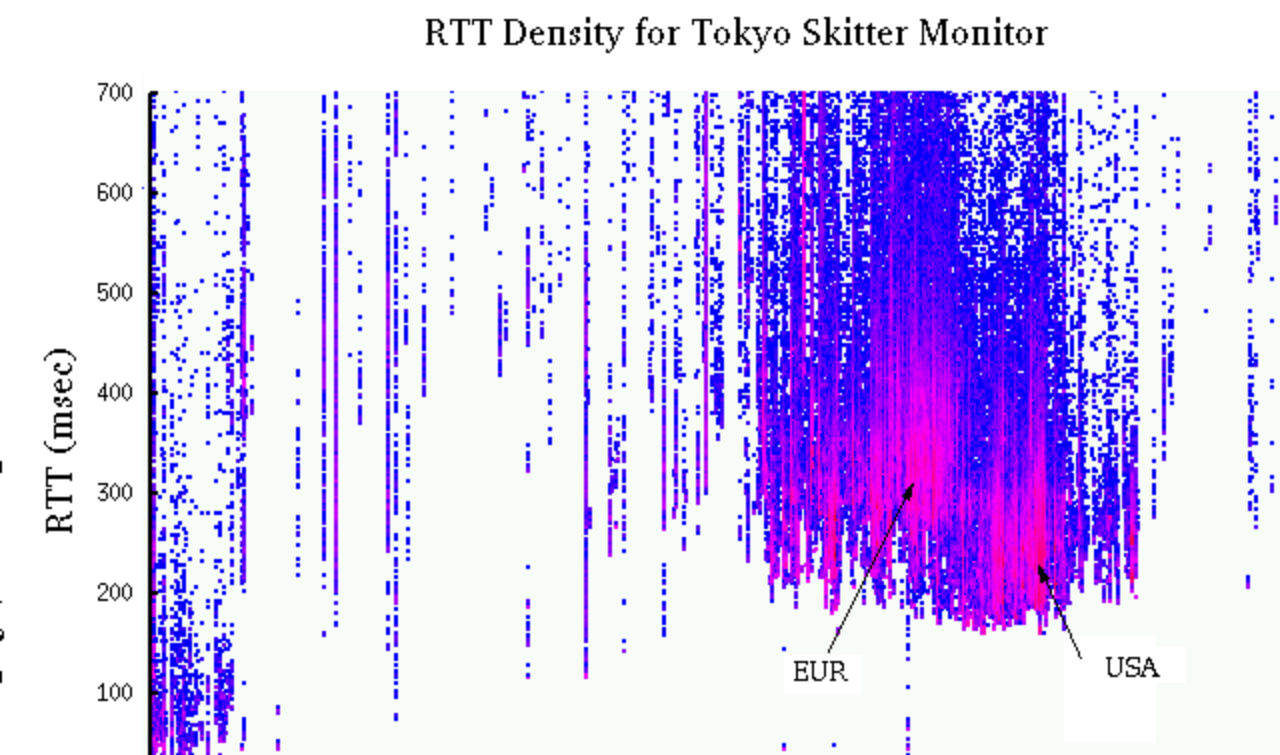
- lower band: directly connected
- upper band: thru US to rest of Europe



skitter: rtt versus distance

tokyo monitor

- european paths ‘close’ but via US



skitter on-going daily summaries

http://www.caida.org/tools/measurement/skitter/summary

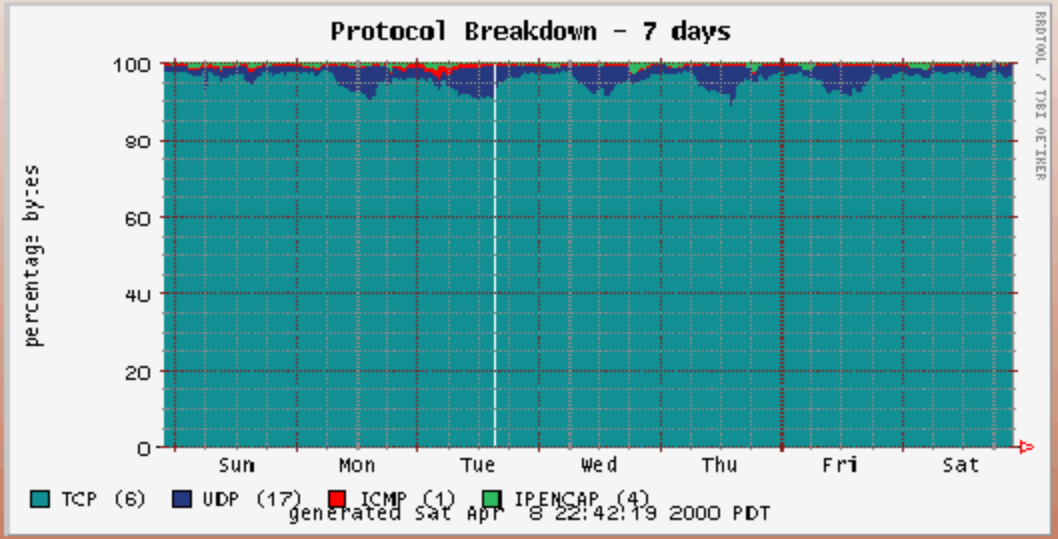
- path length (in IP hops) distribution
- RTT distribution
- RTT versus longitude,
- path dispersion
  - AS & country granularity

Internet workload

- many uses
  - capacity planning
  - performance and QOS assurance across ISPs
  - accounting/billing
  - security management
- measurement tools
  - router–based (cflowd, netflow)
  - stand–alone monitors (coral,skitter)
- visualization huge challenge
  - too much data
  - noone correlates across/with much
- evolution requires use
  - envisioning new methods?
  - better data correlation tools are essential

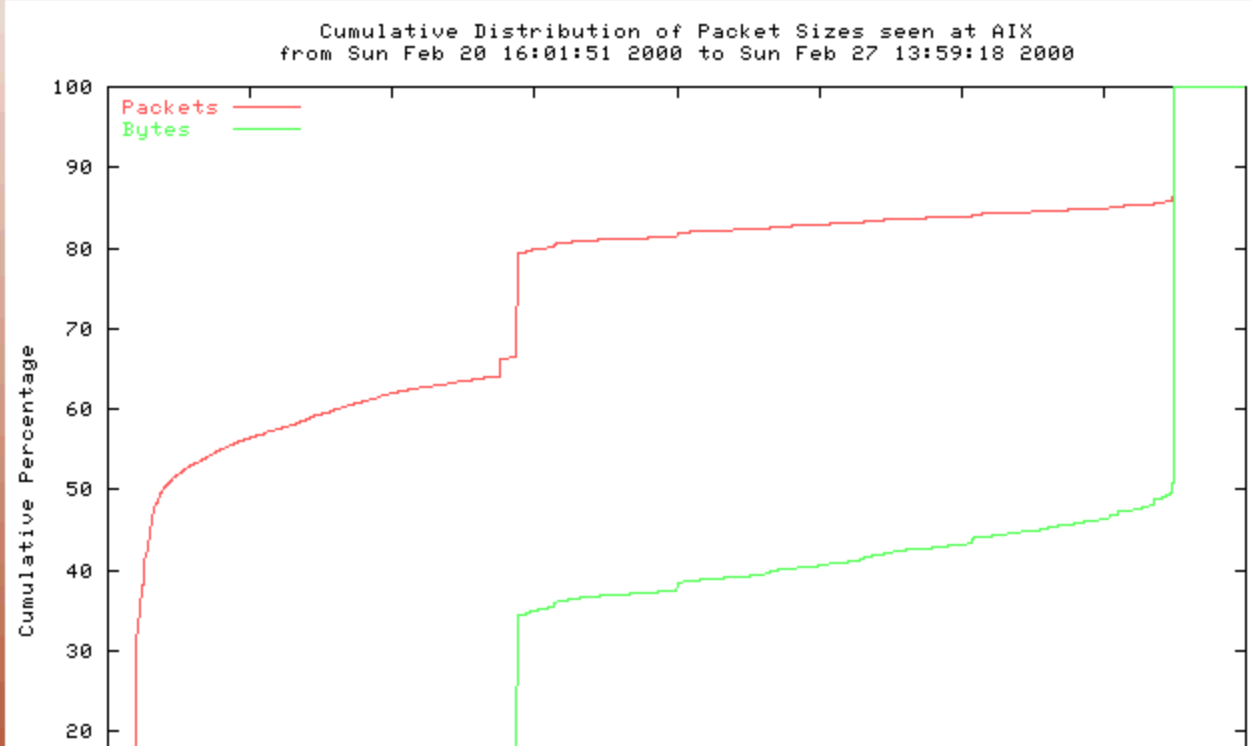
workload by protocol proportions

19 aug 99, ucsd–cerfnet



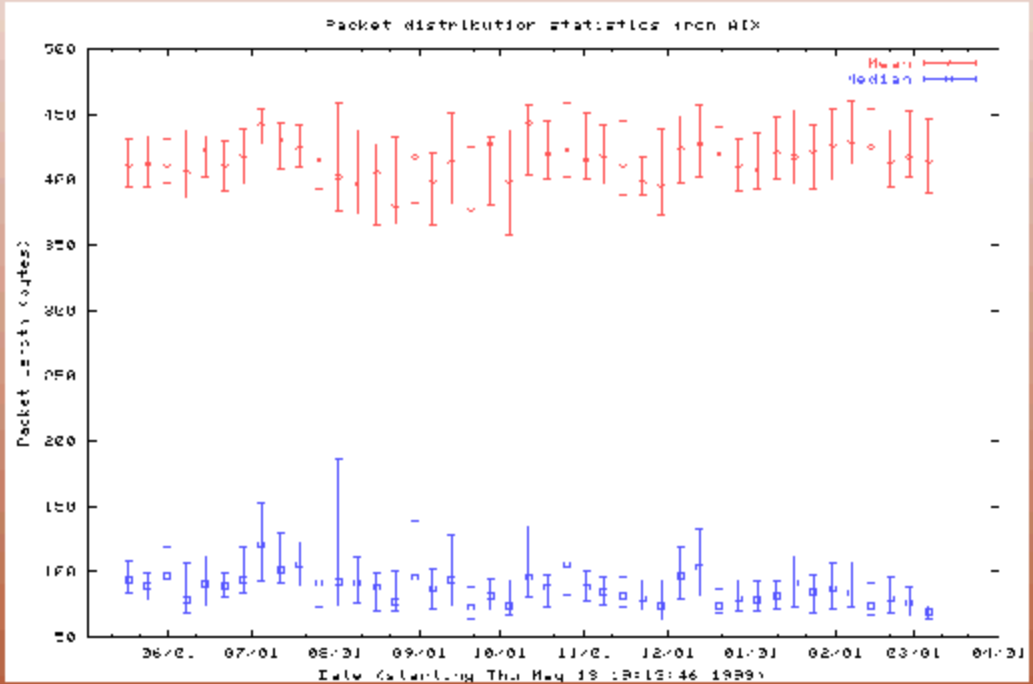
workload: packet size distributions

feb 2000, AMES exchange point
TCP behavior frames distribution



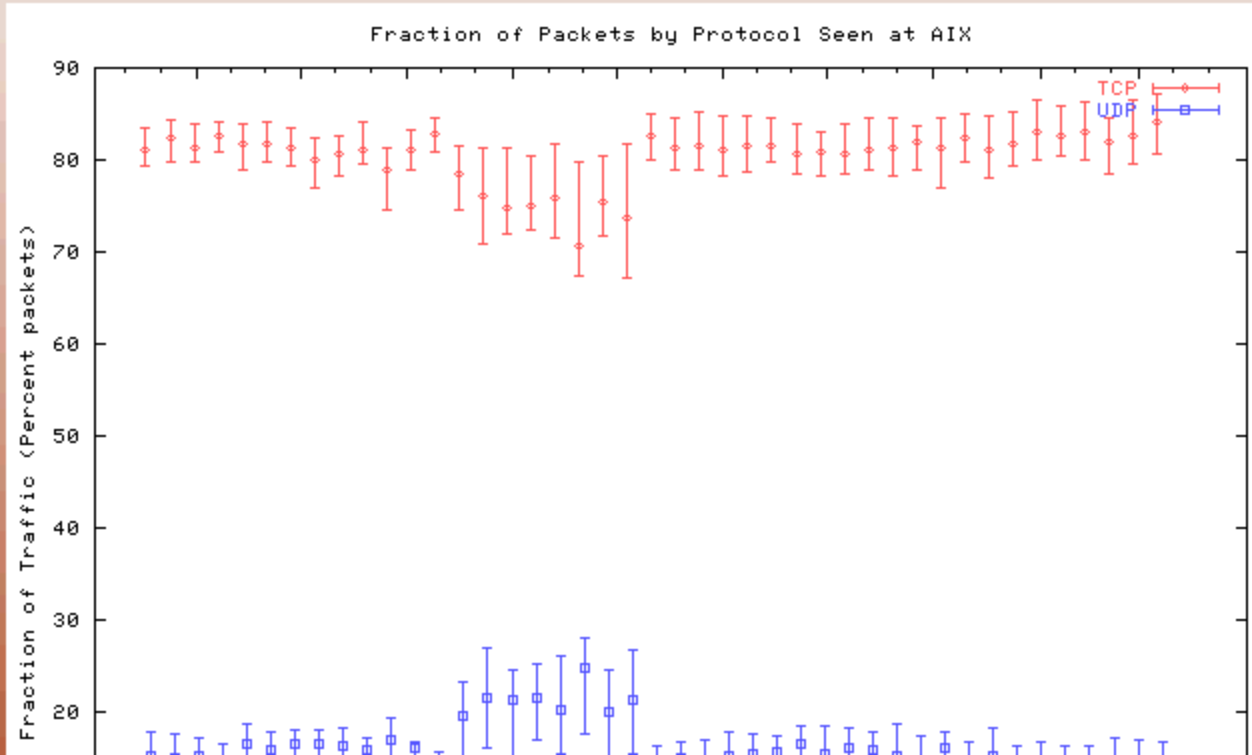
workload: pkt size medians over time

AMES–MAEW xchng pt (1/5 striped links)
little change over 9 months
unsurprising as long as TCP dominates



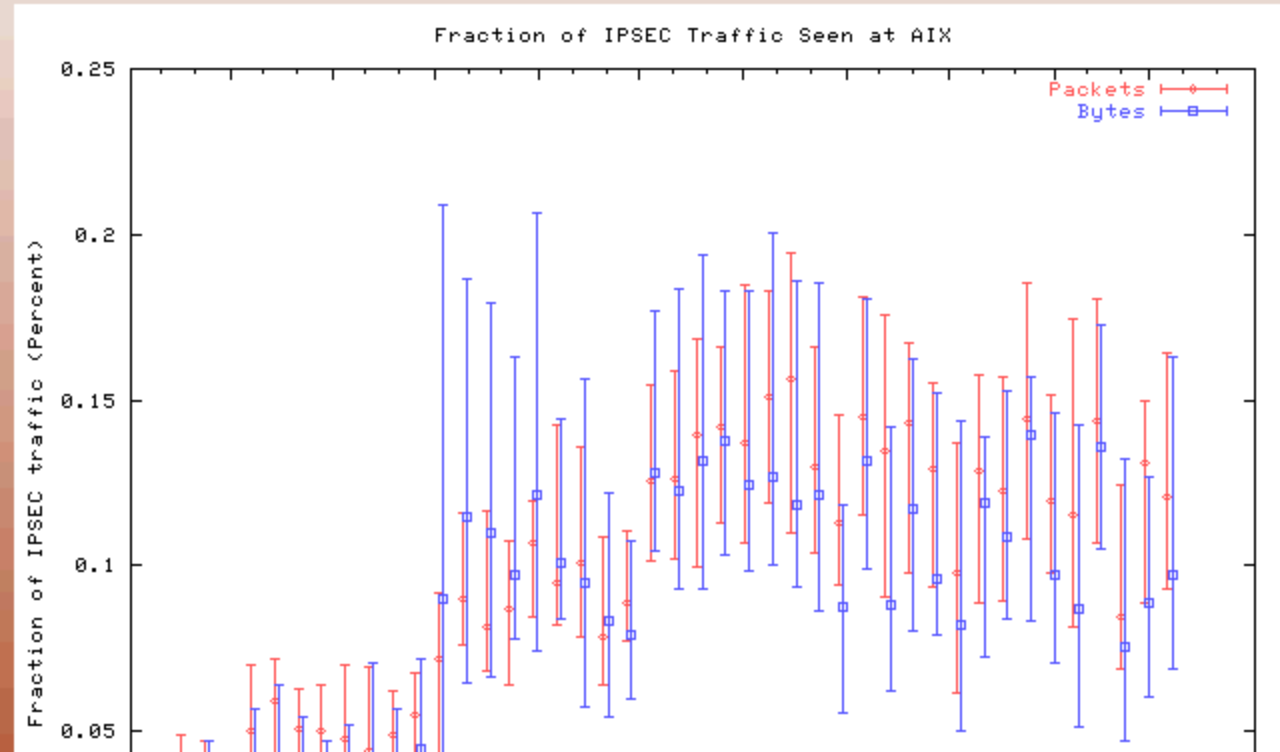
workload: AIX–MAE TCP vs UDP

no clear long–term shift
UDP growth offset by TCP growth



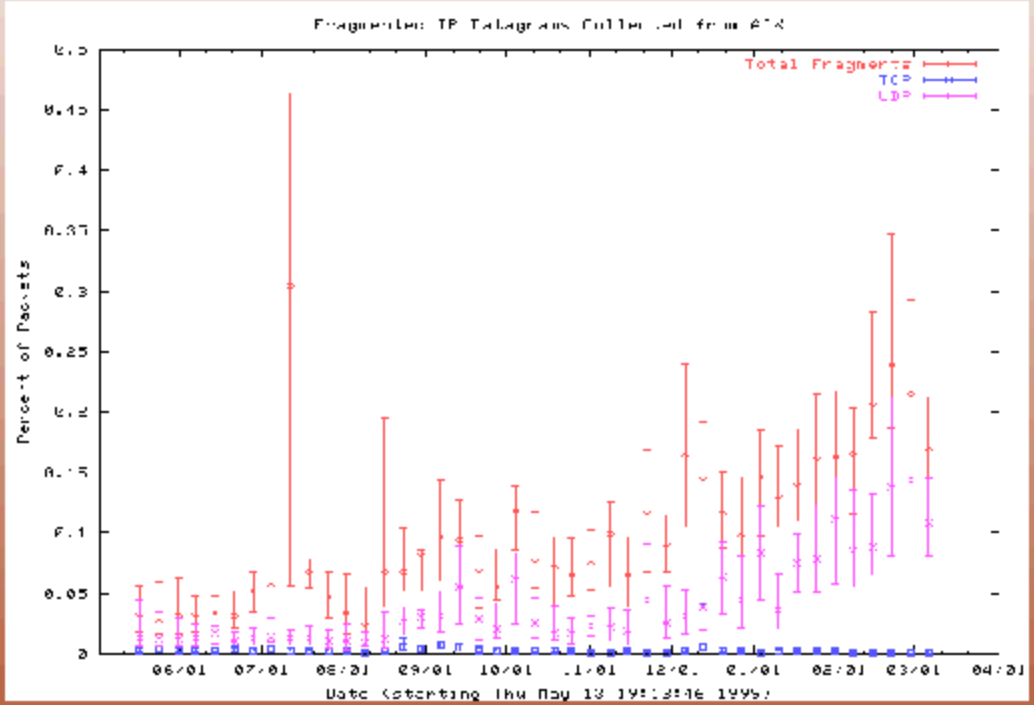
workload: AIX–MAEW ipsec (ah/esp)

almost 10X increase last half of 1999
then levels/declines relative to elsetraffic



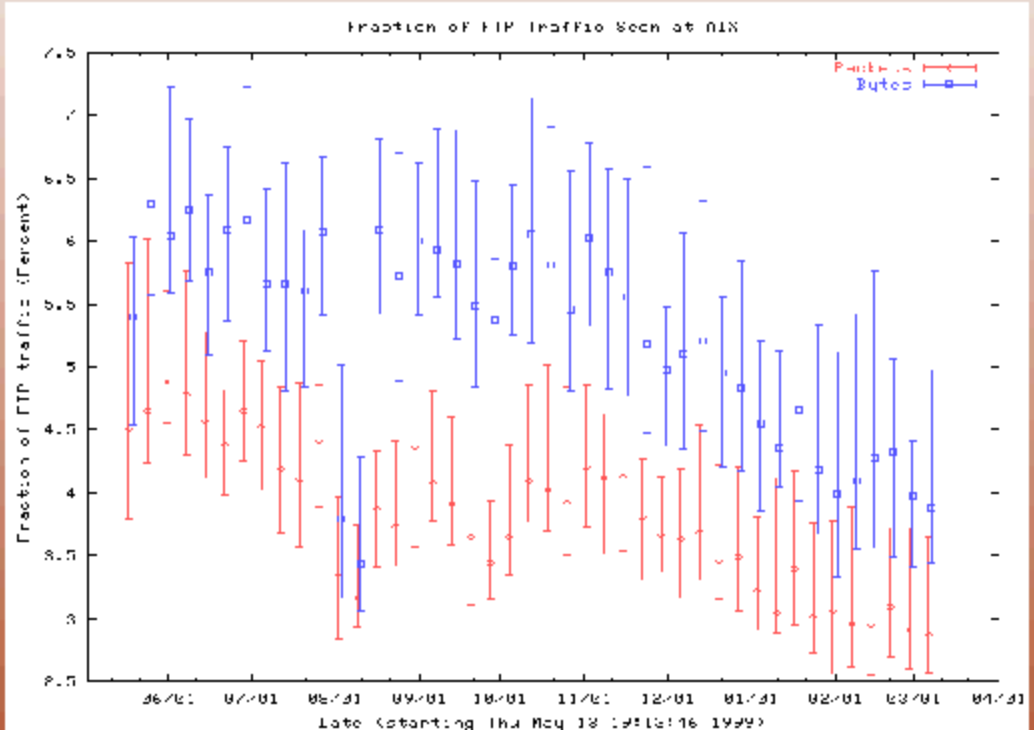
workload: AIX–MAEW fragmentation

relevant to recent IP traceback techniques [Savage00]
definitely on rise (from UDP) at AIX
almost no TCP frags (MTU disc + small pkts)



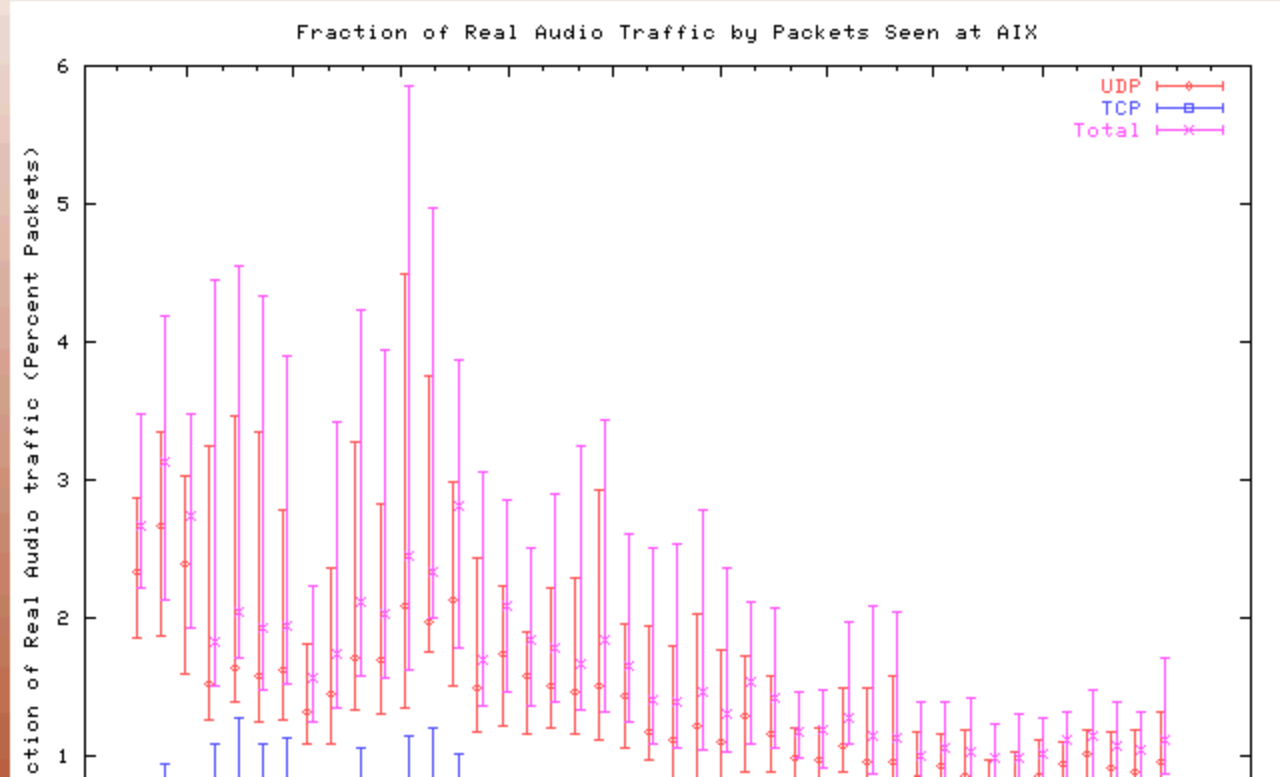
workload: AIX–MAEW ftp

decline in traditional bulk transfer (pre–HTTP)
could be passive mode FTP increase (firewalls)
HTTP doesn't show corresponding % increase



workload: AIX–MAEW realaudio

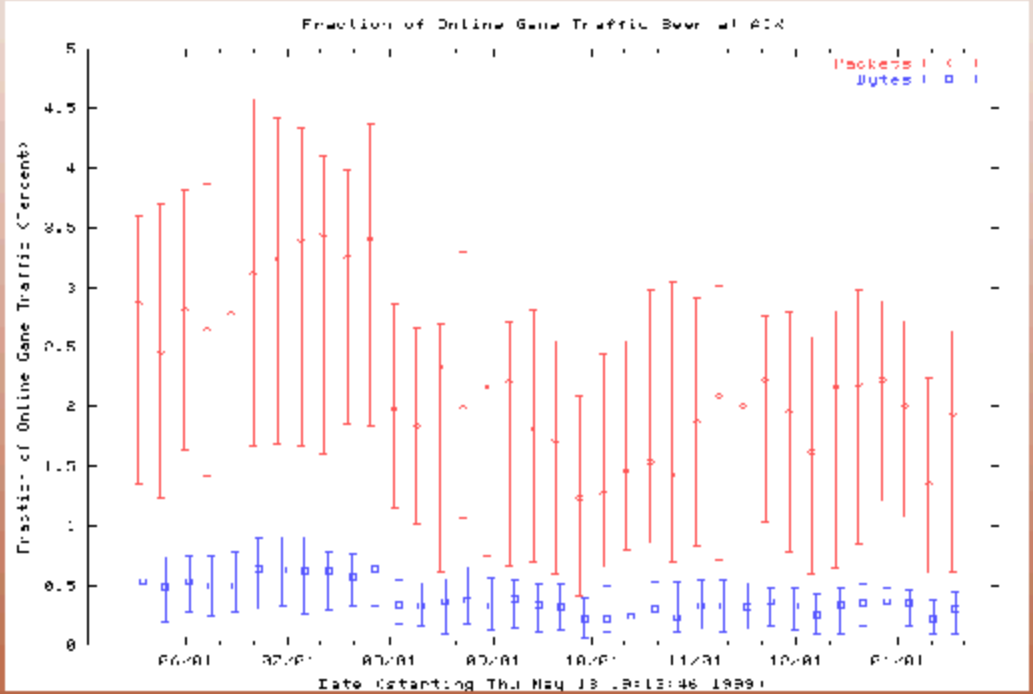
kinda surprising decrease
perhaps other stuff just growing faster (e.g., napster)



workload: AIX–MAEW online game traffic

decline for games included

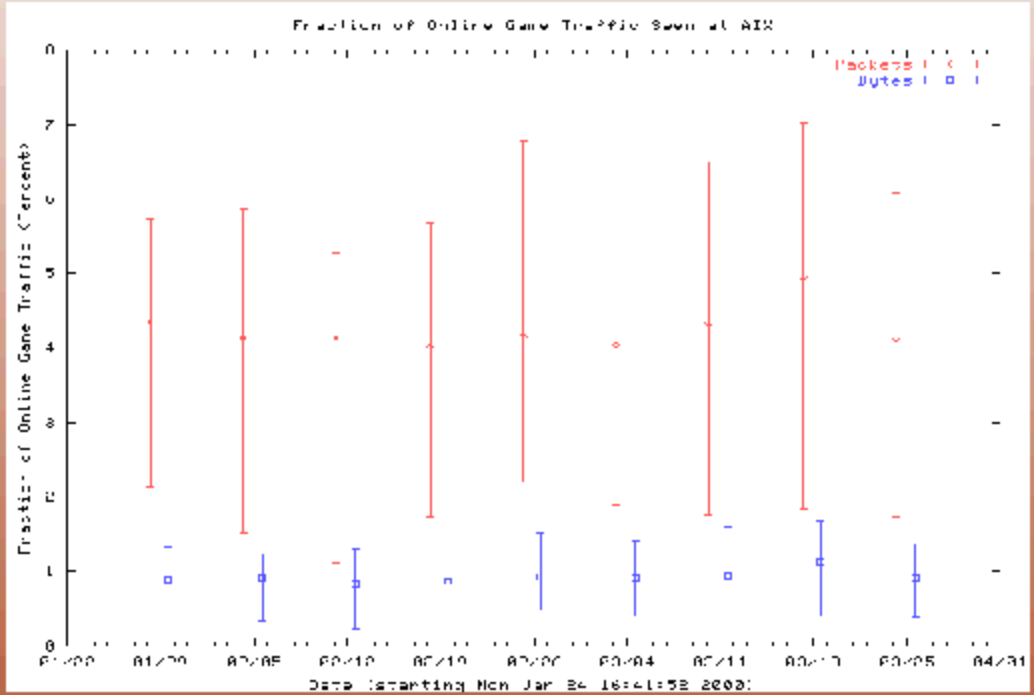
Starcraft, Quake II, and QuakeWorld (a variant of Quake II) popular last year, then other games take over?



workload: online gaming cont.

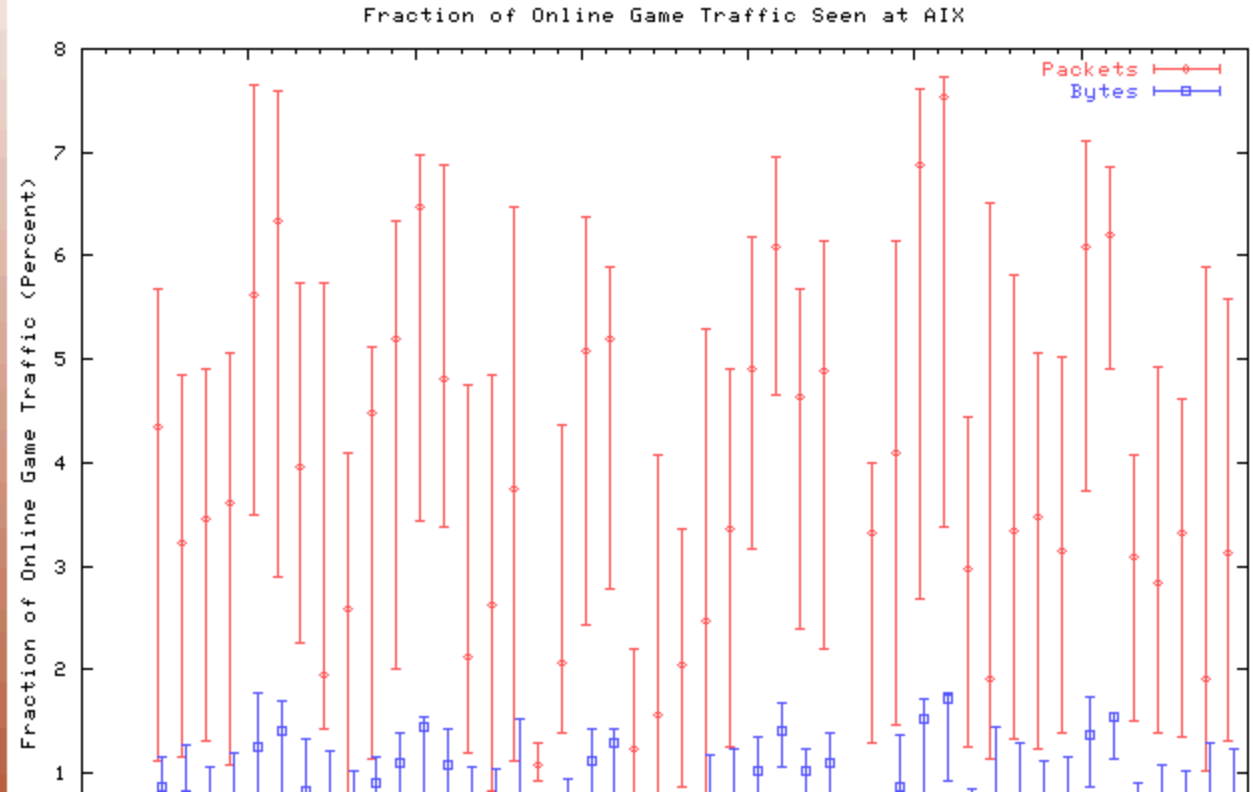
looking at new games plus old

- Half Life, Quake 3: Arena, and Unreal
- median higher –> gaming traffic on rise, but moving target
- increase mostly from new games, older games wane



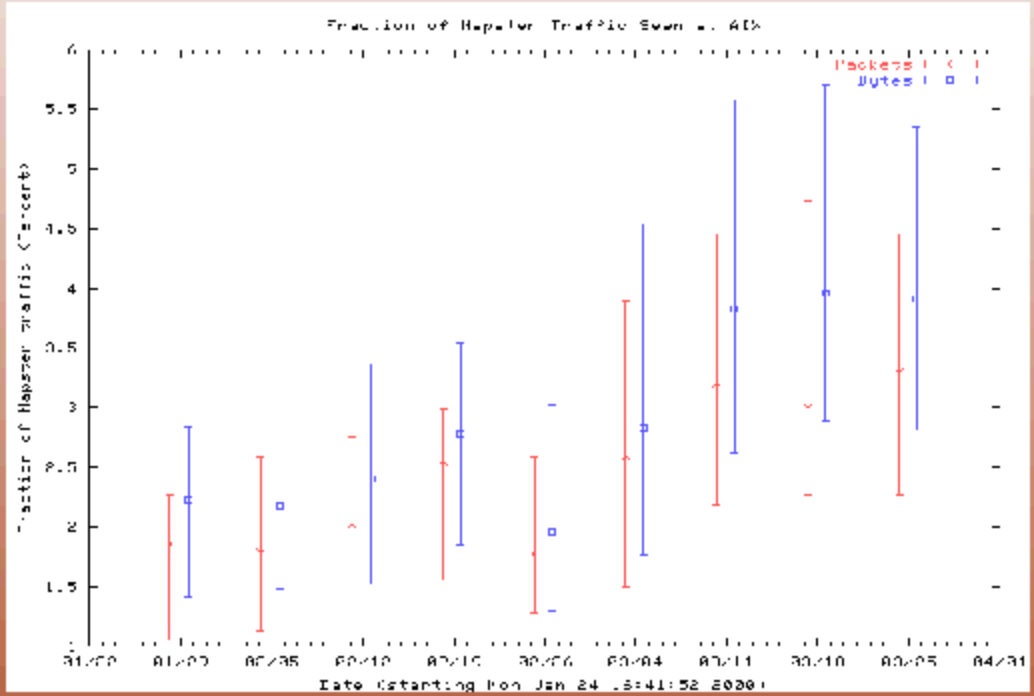
workload: AIX–MAEW gaming trends

clearly more popular on weekends (nearly double!)



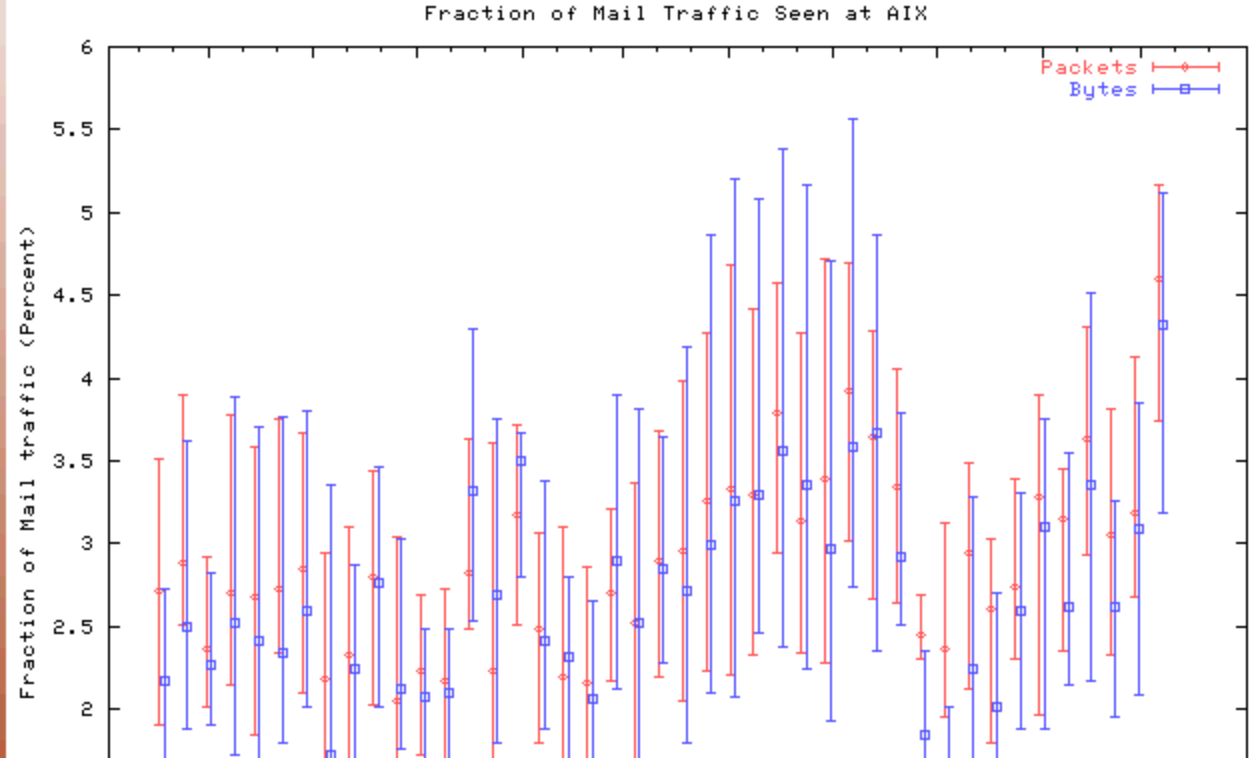
workload: AIX–MAEW napster traffic

3 ports used in late jan: 6688, 6697, 6699
ports migrated in march as universities blocked
even so, dramatic increase (> 50% in feb–>mar)



workload: AIX–MAEW email trends

significant increase around nov/dec, then drops off
online commerce / holiday shopping?



workload: summary of findings

- packet size distribution stable
- TCP:UDP ratio fairly stable
- order magnitude increase in IPSEC mid–last year, then level/decline
- significant increase in UDP fragments
- decrease in active–mode FTP and realaudio
- increase in gaming and napster
- increase in email during holidays
- strong weekday/weekend pattern in gaming

workload: meta-challenges

- splintered & competitive core
  - limited access to data
  - so difficult to argue 'representativeness'
- network performance impact
  - higher b/w increasing difficult to measure
  - faster speeds and changing transport technologies complicate data acquisition and processing
  - e.g. monitor gone when AIX converts to POS
- user privacy volatile issue
  - hard to get data in researchers hands

CAIDA's UCSD/CERFnet link monitor available:
https://anala.caida.org/CoralReef/Demos/

workload: challenges

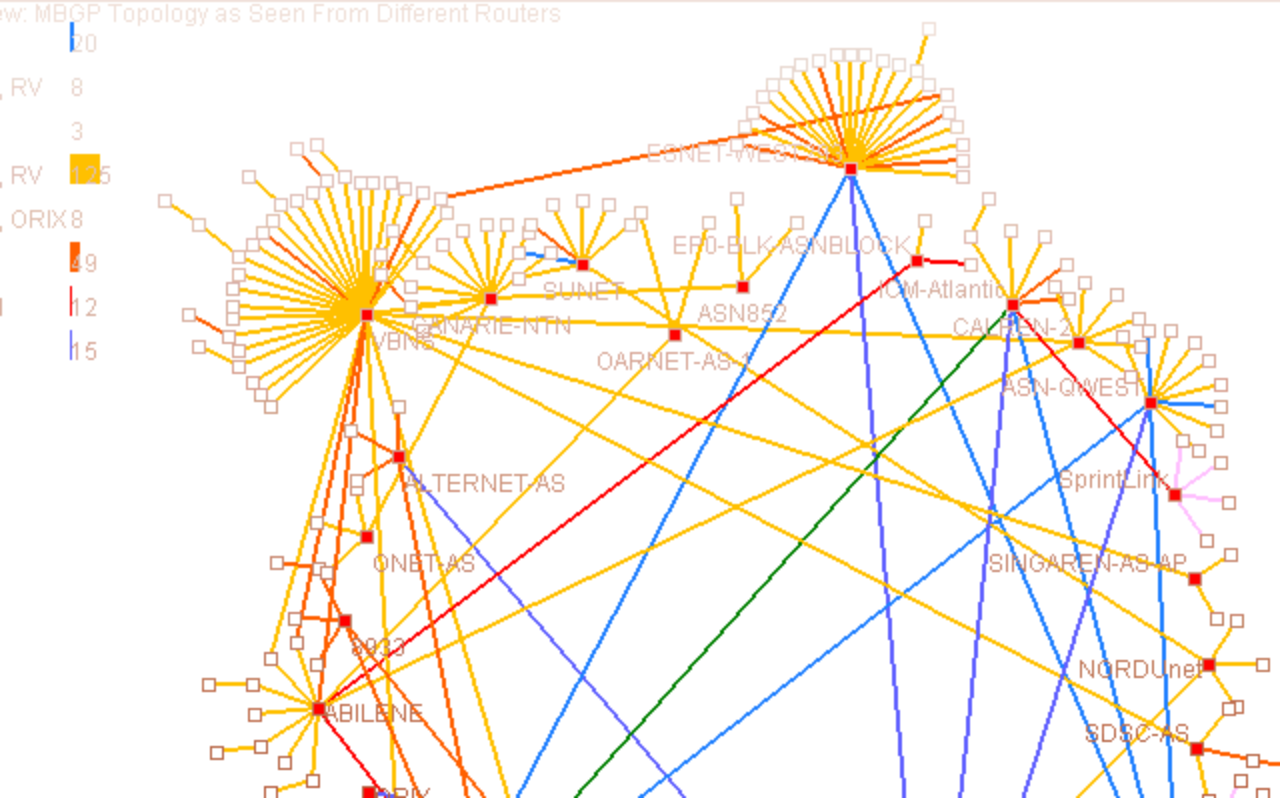
- id and present 'useful' workload metrics, particularly given persistence of fire-fighting environment
- id significant patterns, timeframes, correlations
  - vary by user need
  - change as technologies and 'net change
- methodology has many weaknesses
  - dynamic port negotiation (napster)
  - tons of 'other' ports unmapped
  - ports not really assurance/unique anyway
  - IPSEC blows away ports anyway
  - need traffic profiling
  - things getting worse not better here

routing & addressing

- currently non-intuitive, incantation-driven
  - configuring is close to black art
  - sausage/laws....
- real-time identification of outages, flaps, critical paths
- routers really won't give you much data
  - "forwarding or stats: pick one" -tli, paraphrased

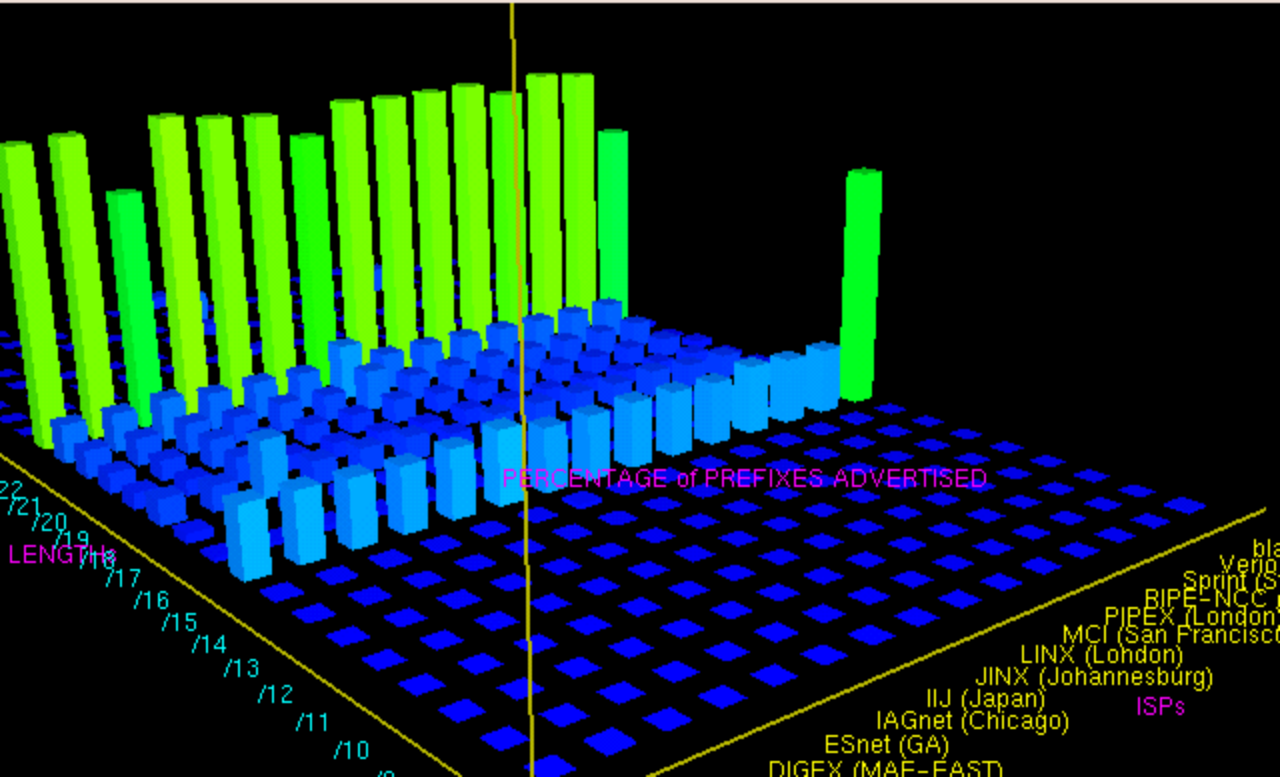
topology: differencing routing tables

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



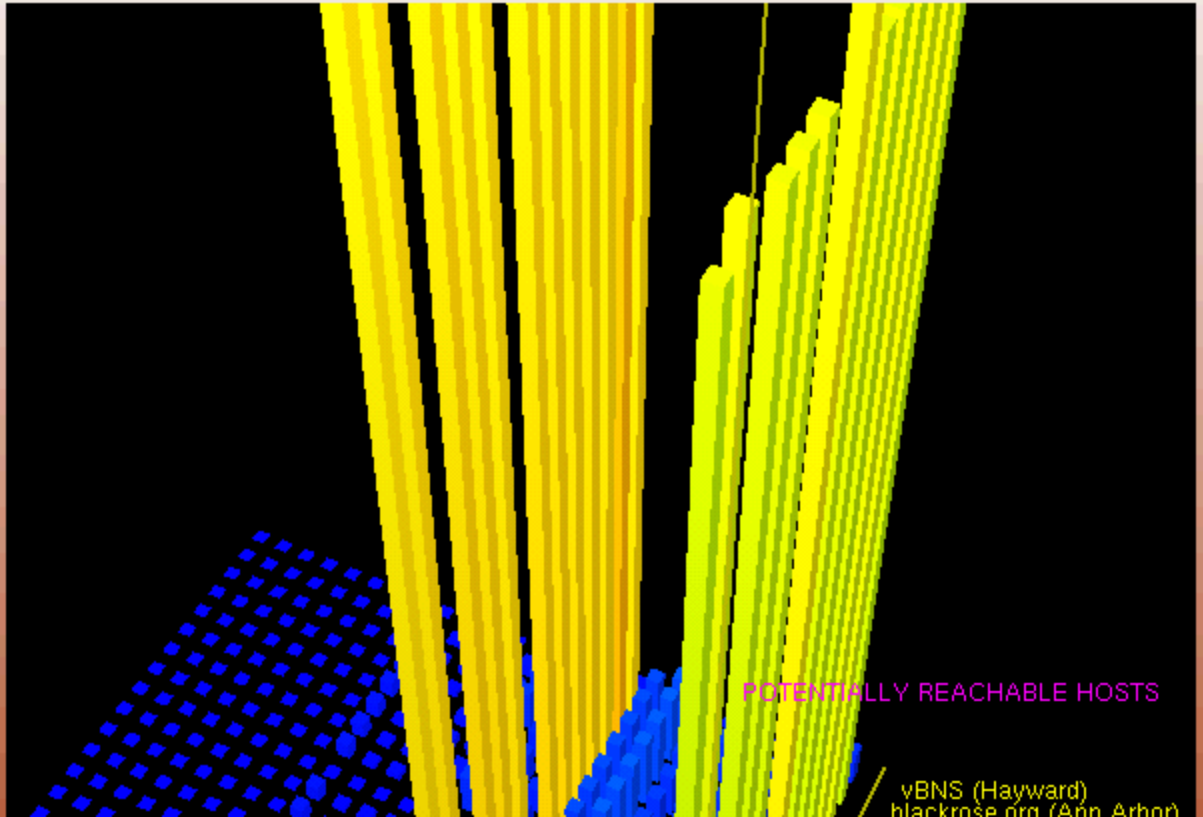
routing: address consumption

prefix length distribution for routes announced by core ISPs, 1-6/1998 (courtesy NLANR/MOAT, Jeff Brown)



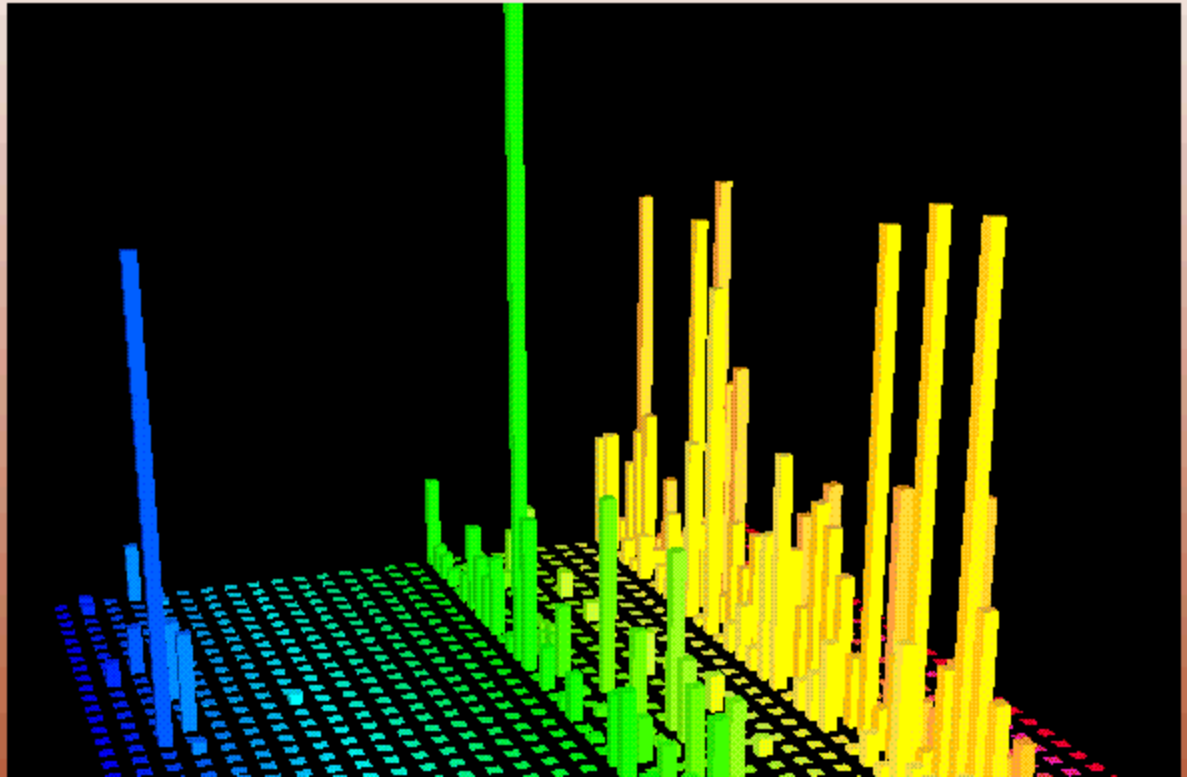
routing: address consumption (#hosts)

reachable hosts for routes announced by ISPs, 1-6/1998



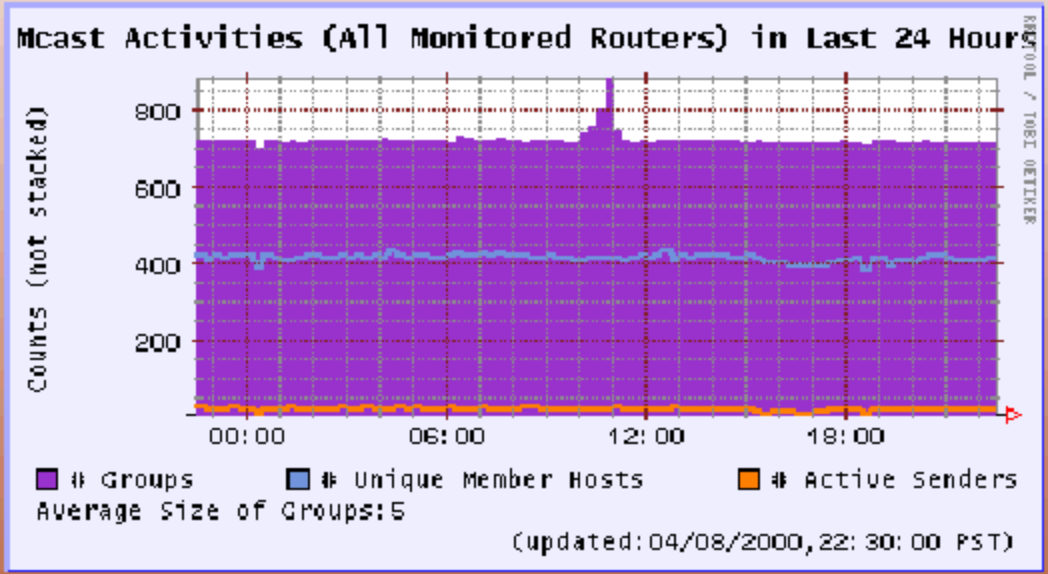
routing: address usage of *traffic* sample

32x32 'bitmap' matrix of address space
height is % packets with src IP in that address block



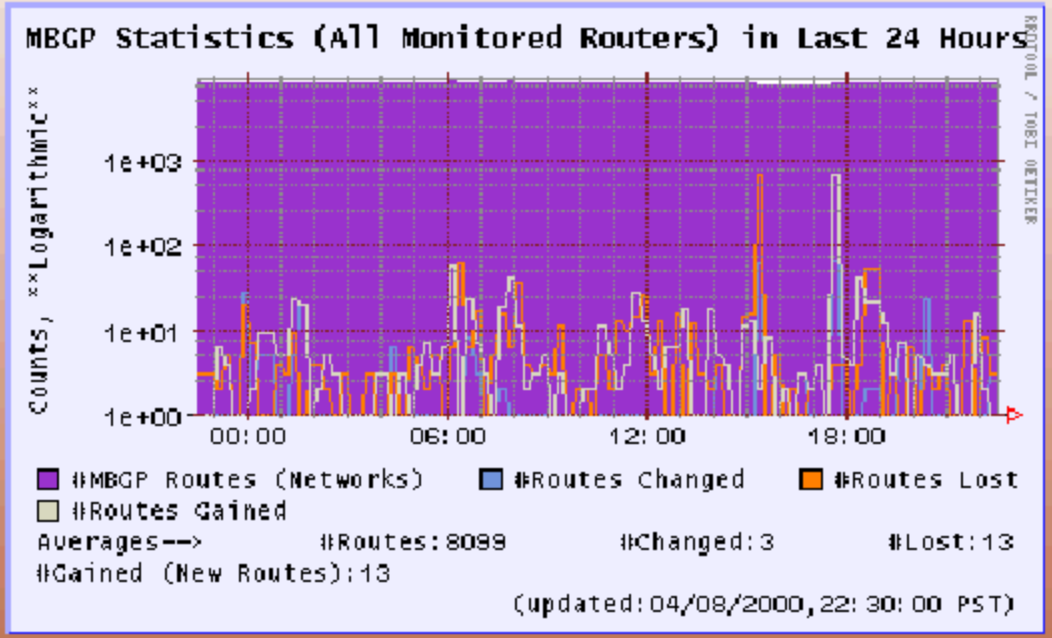
multicast workloads (using mantra)

8 apr 00, fix-west.mbone.nasa.gov
daily updates



multicast (using mantra)

daily updates @ http://www.caida.org/Tools/Mantra/
mbgp, msdp statistics, topology maps/diffs



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 performance problems with some measure of path 'length'
  - comparison of forward path with
    - BGP path
    - shortest path
  - does asymmetry matter?
  - effects of unicast/multicast incongruities?

routing vis: research obstacles

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

now what?

- the ideal:
  - well-instrumented infrastructure
  - seamless integration of variety of data sources
  - important for simulation/prediction
  - but unlikely for the foreseeable future
- tools still need:
  - interpret of vast quantities of data in real-time
    - geographically & logically distributed
  - user-friendly integration with network utilities and control systems
  - inter- & intra-ISP feature detection
  - new methods for data collection, reduction, aggregation, and mining (GByte or Tbyte datasets)

www.caida.org/Presentations/

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