measuring the 'real' Internet:

acquisition of empirical data in support of Internet modeling and simulation

19 jul 2000
darpa nms kickoff

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

outline: data sets for NMS research

- where we are
- how we got here
- what data we have now
- what data we can get soon
- what we should do next

Internet's resistance to modeling

evolution–based (good!) reasons

- protocols, technologies, applications
  - independently developed and deployed
  - by no means syngergistic
  - by all accounts rapid
  - 'punctuated' but no equilbrium

but simulation/analysis validation requires core data

- right granularities hard to come by
- measurement technology just not there
- argument for it also not there
- losing battle?

Internet's resistance to measurement

- many would benefit
  - vendors, users, researchers, ISPs
- ISPs would bear cost
  - multiple media: atm, pos, dwdm, mpls
  - logistics/management
  - privacy implications
  - analysis/research obsolete after (before) done

.....how to justify measurement??

one answer: tools ISPs want or need

measurement tools lack

- well–defined traffic metrics
  - e.g supporting SLAs or billing
- uniformly applied methodologies
  - varied topologies, equipment, ISP practices
- clear definition of measurement hypotheses or goals
- measurement scalability
- ability to explain phenomena
  - topology changes, routing loops, black holes
- relevance to actual ISP problems or mechanisms for fixing
- communication of useful results

publically available data sets

- not comprehensive
- no such thing as representative
- case study mentality behooves us

- data sources, tools
- student–compatible

Internet measurement taxonomy

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

correlation essentially non-explored
.....(holistic Internet measurement?)

Internet topology data

why do we care?

- simluation and modeling validation
- traffic engineering
- track global growth/change
  - arguable: increased potential/manageability

macroscopic, IP layer

- Lucent: burch/cheswick maps
- CAIDA:

  www.caida.org/tools/measurement/skitter

topology: caida's skitter

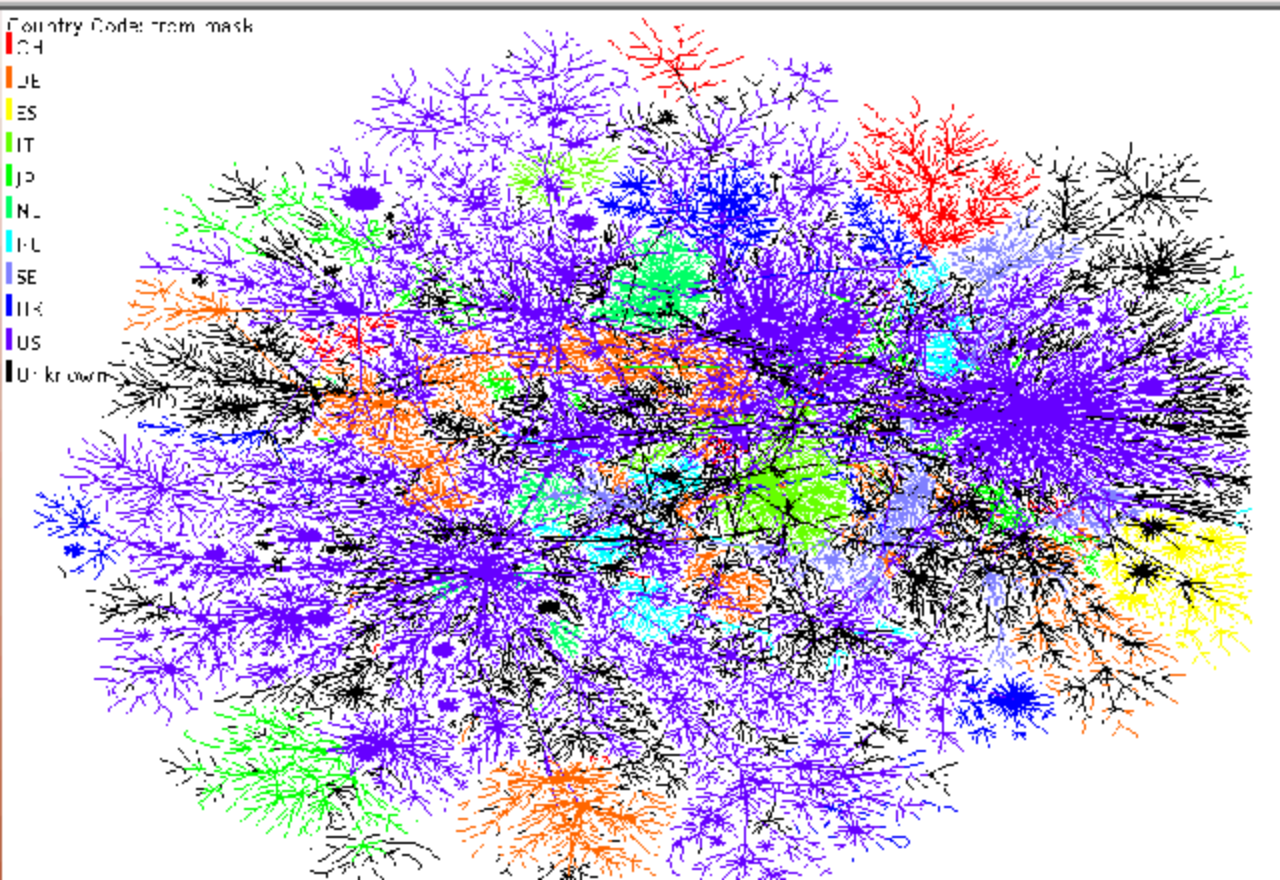
- track/depict topology cross-sections
  - 22 monitors (inc. some root name servers)
  - forward IP path and round-trip delay
  - tens of thousands of dst (mult. lists)
  - remove targets that complain
  - 150k nodes, 270k links in 5 days (11/99)
- architecture
  - continuous, parallel 52-byte ICMP probes
  - depending on dst list size, O(1) probes/hr/dst
  - kernel time stamping
- correlate path perf. w events, e.g. BGP
- identify critical pieces of infrastructure
- case studies of relevant cross-sections

other active (probed) data sets

- MOAT: http://amp.nlanr.net
  - HPC sites
  - RTT, traceroutes
- I2's Surveyor:
  http://www.advanced.org/surveyor/
  - I2 sites
  - one-way delay, paths
- vBNS http://www.vbns.net:8080/stats/
- SLAC, NIMI, others

http://www.caida.org/tools/taxonomy/
http://atlas.caida.org/

skitter: colored by countries



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

topology/perf.: priorities

- dynamic feature detection from large, complex datasets
  - data aggregation/reduction techniques
  - faster data collection, processing, rendering
  - meaningful displays, user-friendly tools
  - correlation with different datasets
- large scale public database of performance data
  - across many sources
  - comparisons w/topology, workload, routing analyses

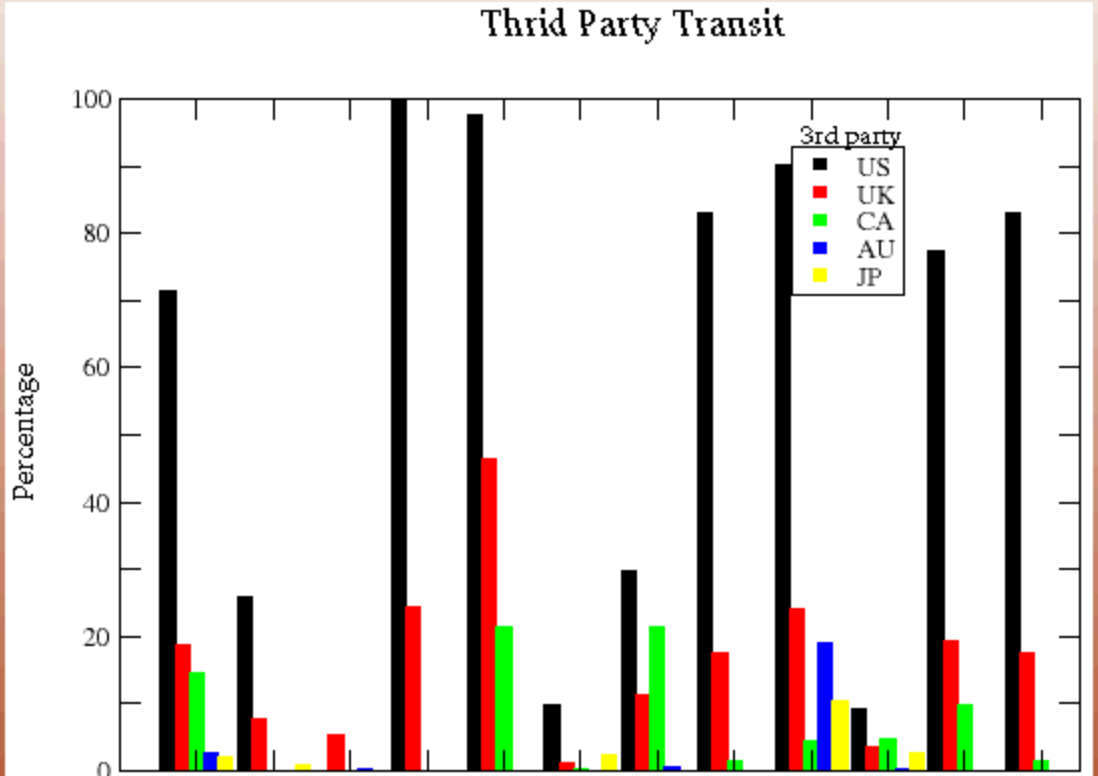
topo./performance: obstacles

- poorly defined user requirements/interfaces
- negative perceptions regarding quality and worth driven by explosive growth
- uniform methodology impossible
- mapping IP addresses to ... anything meaningful
  (not just geography)

.....things getting worse not better

BGP policy: US as transit

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



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)

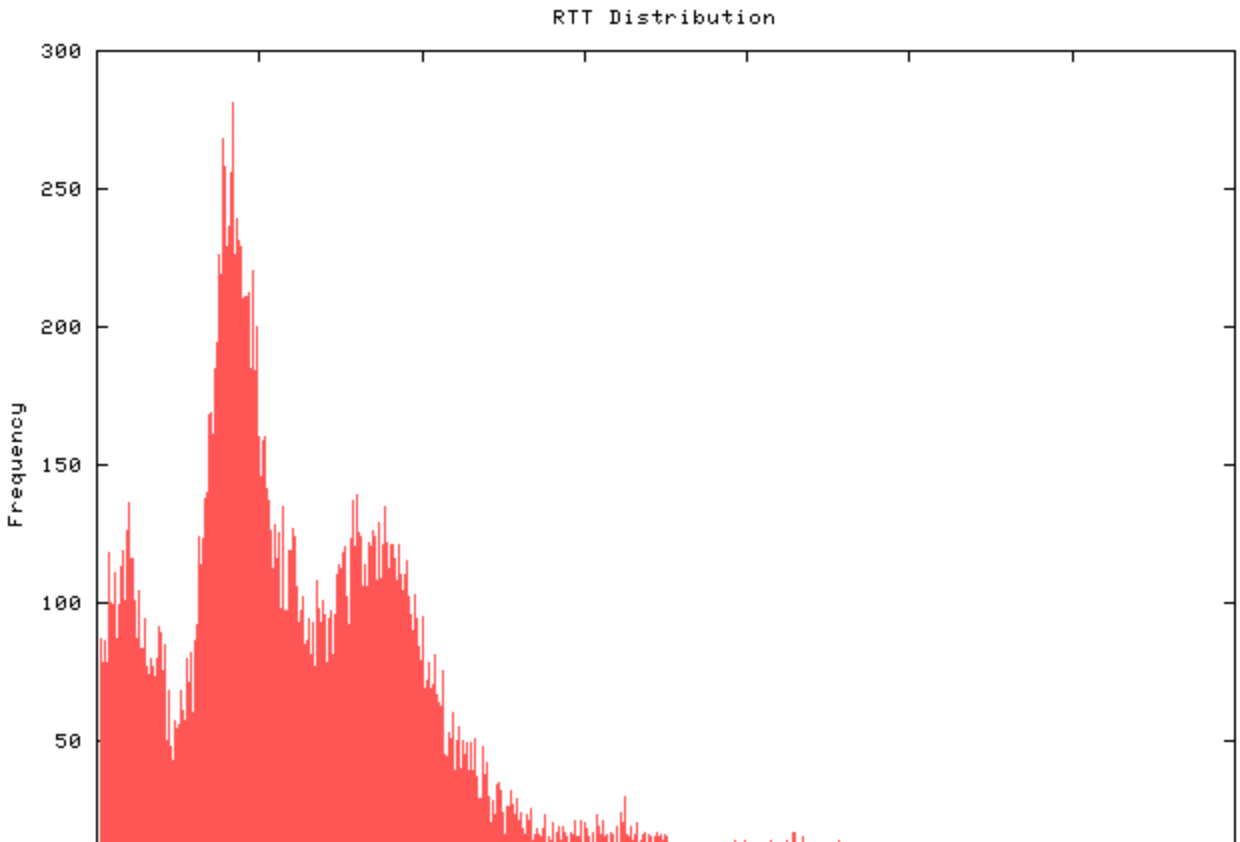
	all	AU	CA	C_H	JP	KR	MX	NZ	SEA	SWA	TW	US
US	71.5	77.8	82.0	90	49.5	61.6	100	79.6	63.0	97.8	83.5	
CA	13.3	8.3			4.9	37.5	2.1			27.5	22.3	1.3	0.2
AU	2.8				18.4					46.1	1.6		
0.4													
JP	1.2			1.4	7.4		10.5			12.0			0.3
NZ	0.9	3.7											
EUR	0.7				2.1		1.7			4.2	27.0		
UK	0.7	0.0	0.0			0.1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

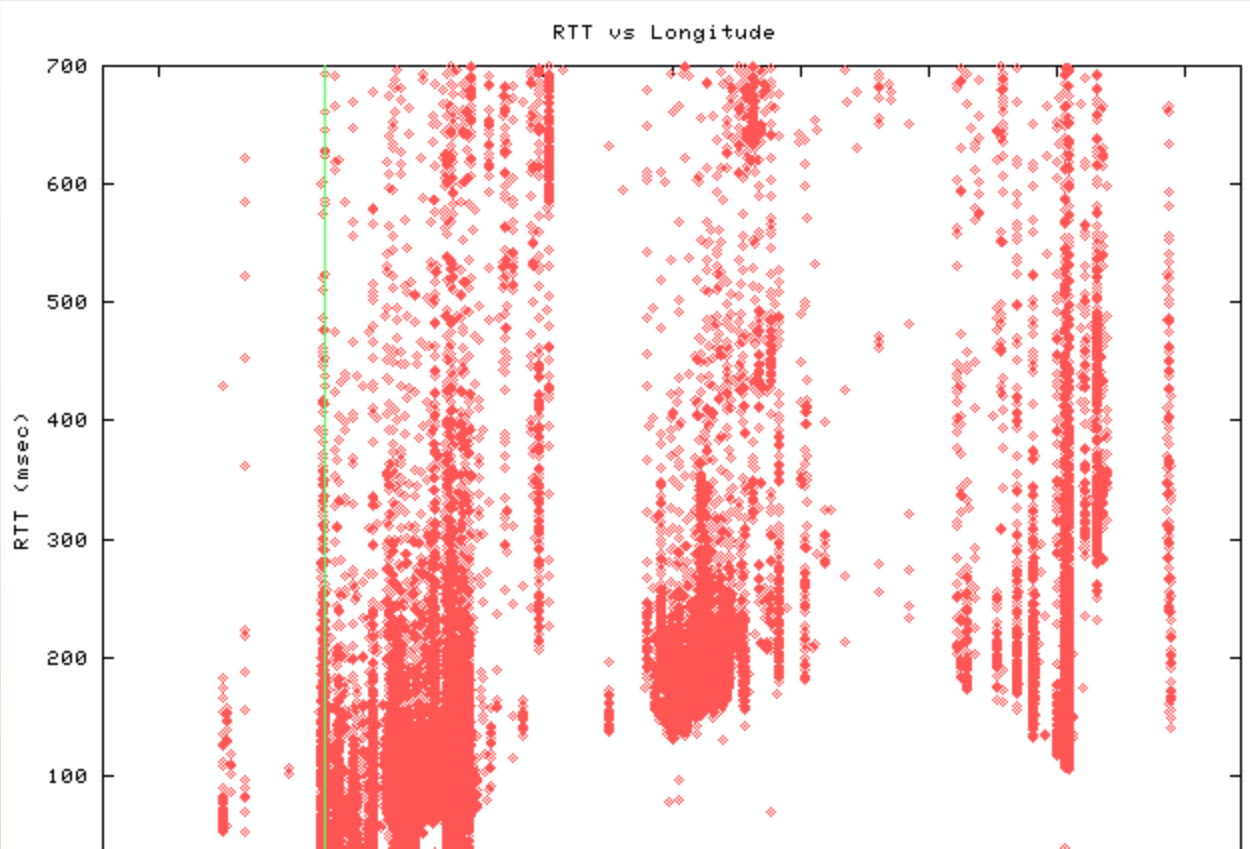
- RSSAC, DNS technical advisory committee to ICANN
- goal: optimize root nameserver location
  - co-locate skitter hosts w root servers
  - demonstrate root server performance in serving target community
  - develop techniques for evaluating architectual optimality for root server placement
  - visualization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

(www.caida.org/tools/measurement/skitter/)

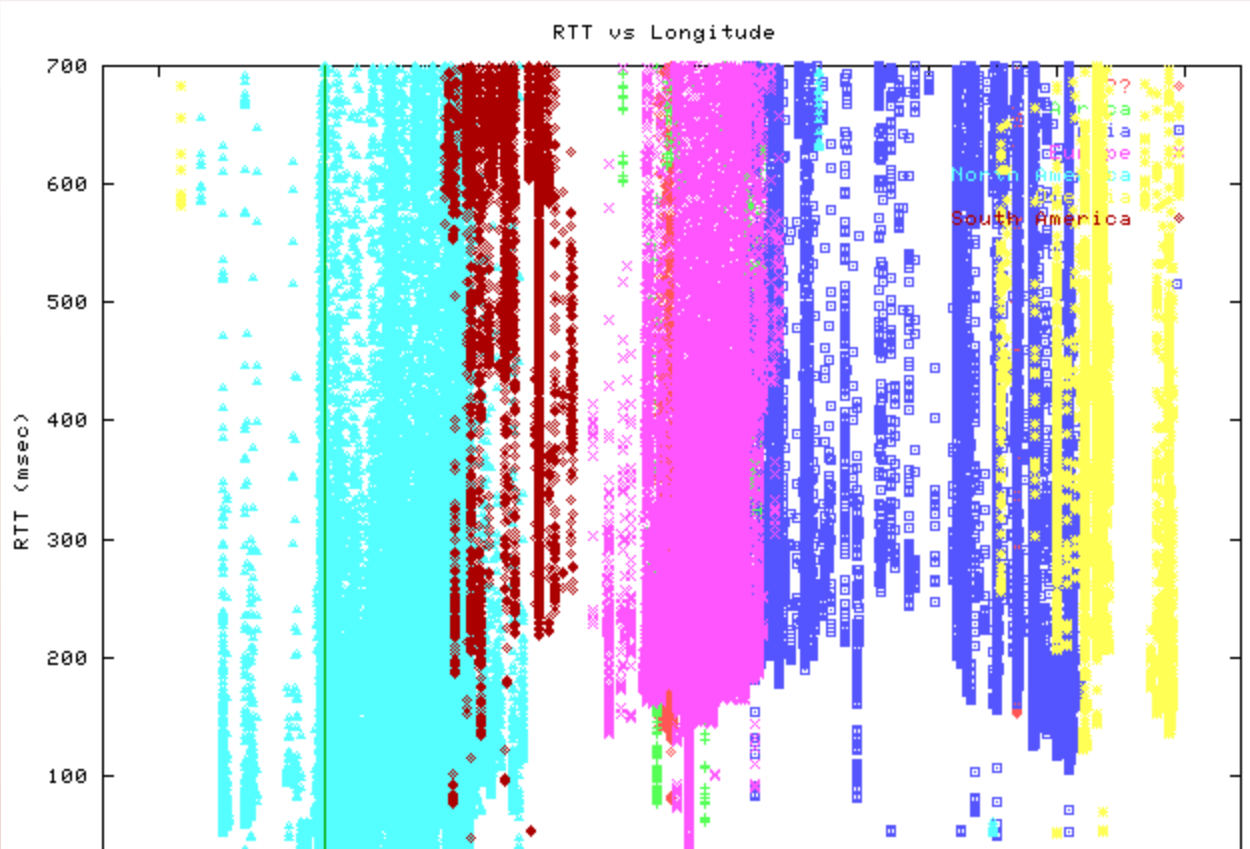
skitter: rtt distribution: tri-modal



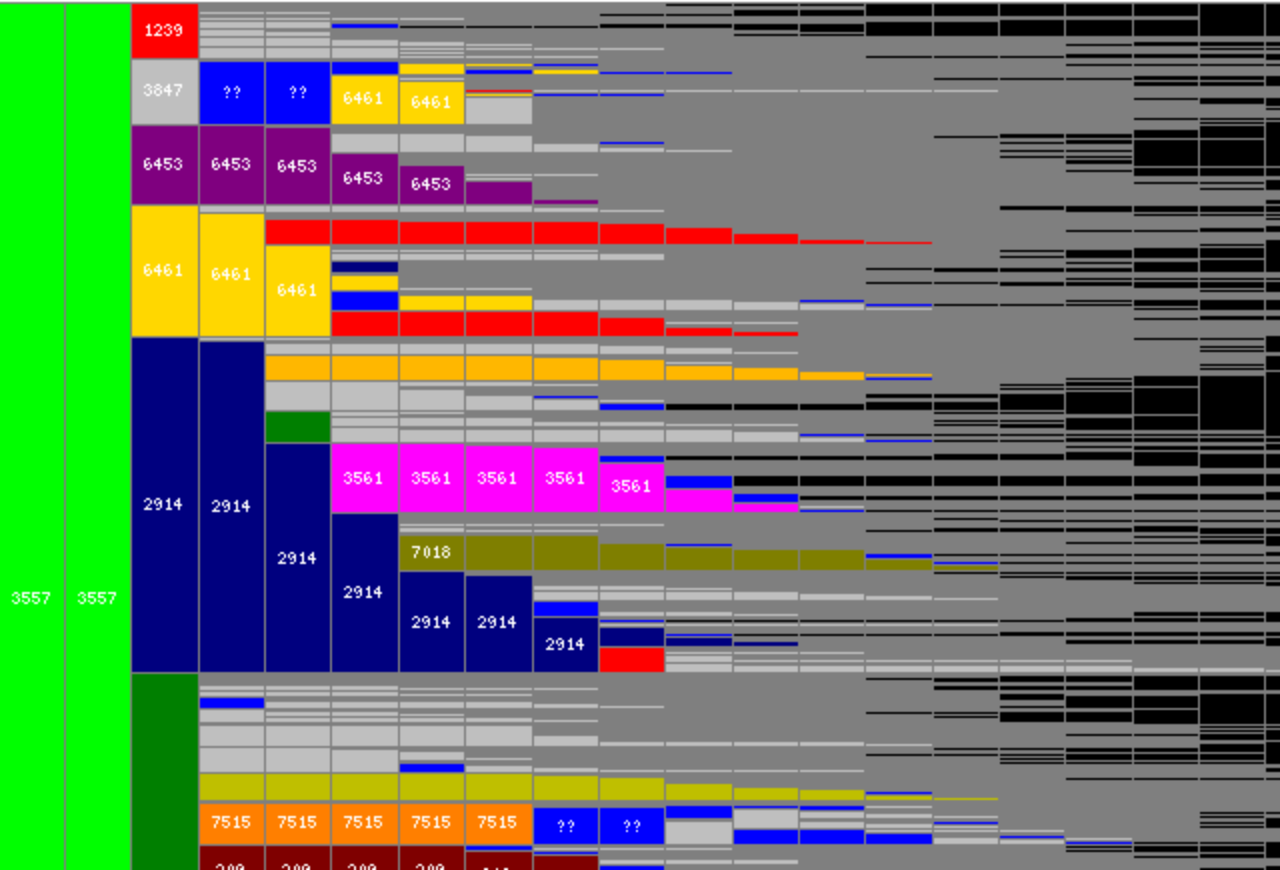
skitter: rtt vs longitude (light cone)



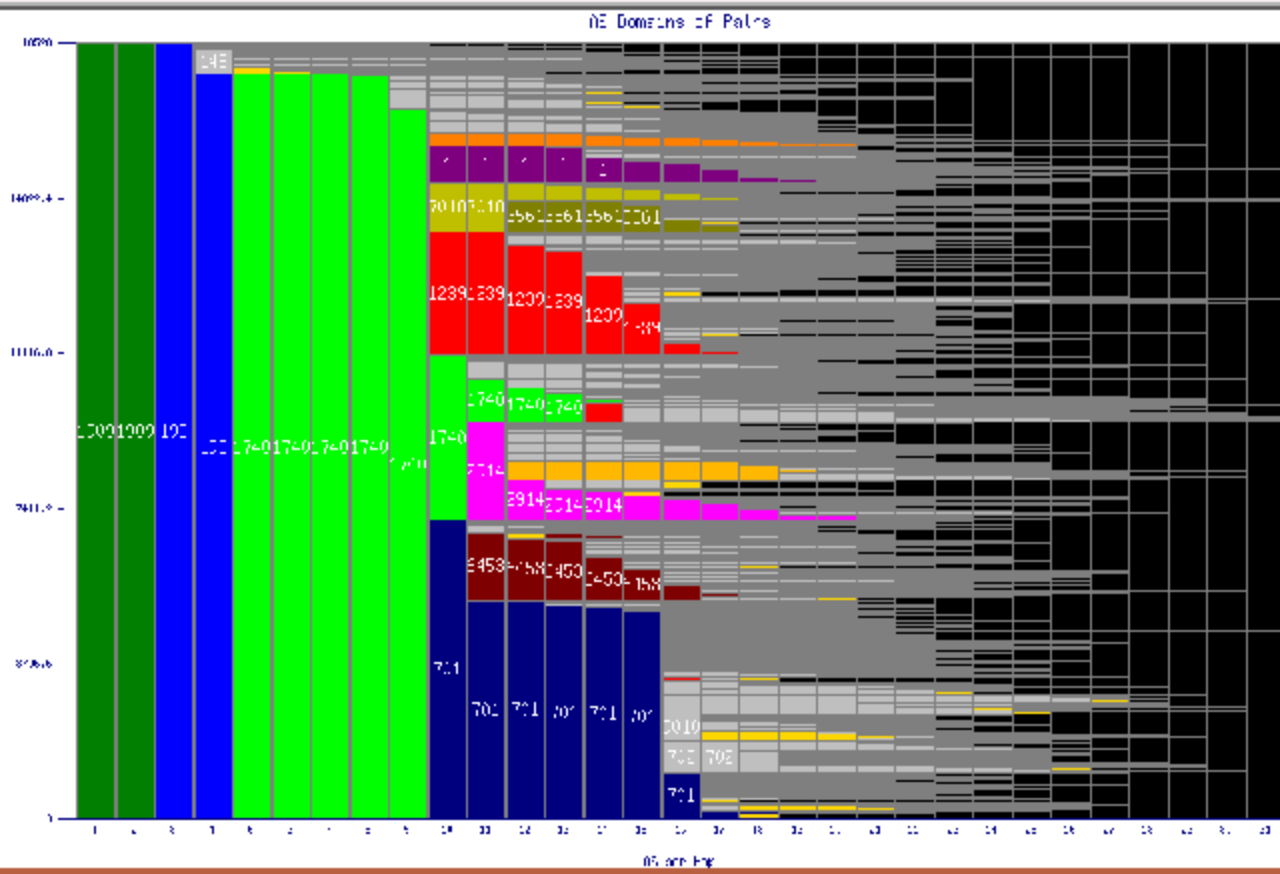
skitter: rtt vs longitude (light cone)



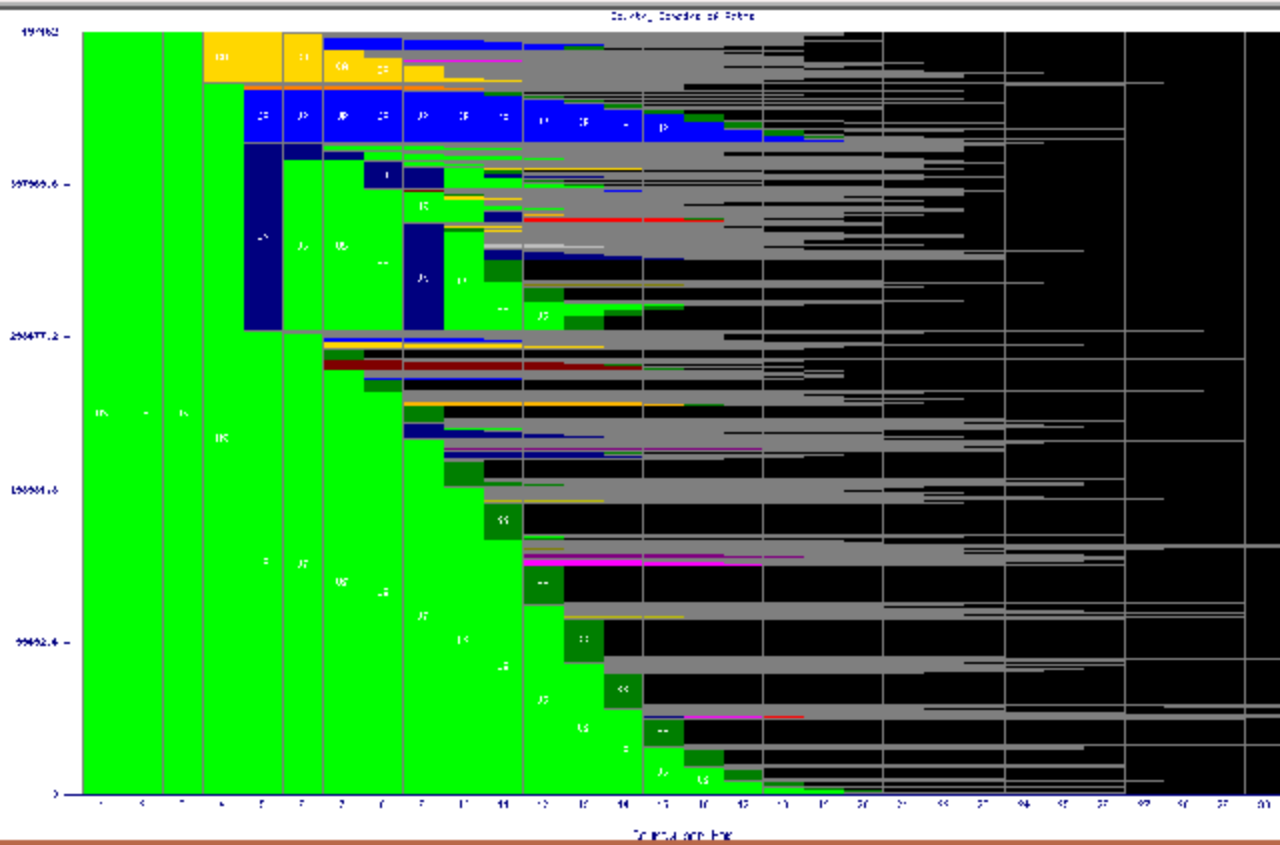
skitter: dispersion among ASes across paths



skitter: AS dispersion across paths (sdsc)



skitter: country dispersion across paths



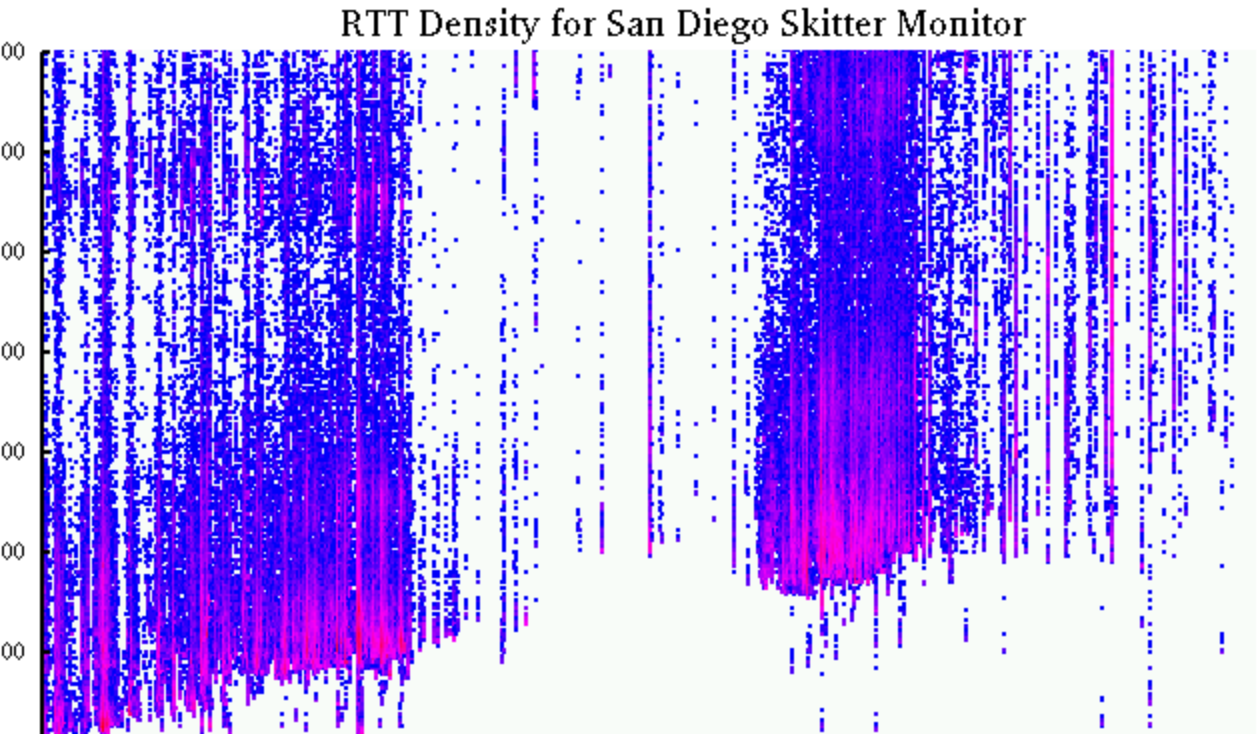
DNS roots study: future

- 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

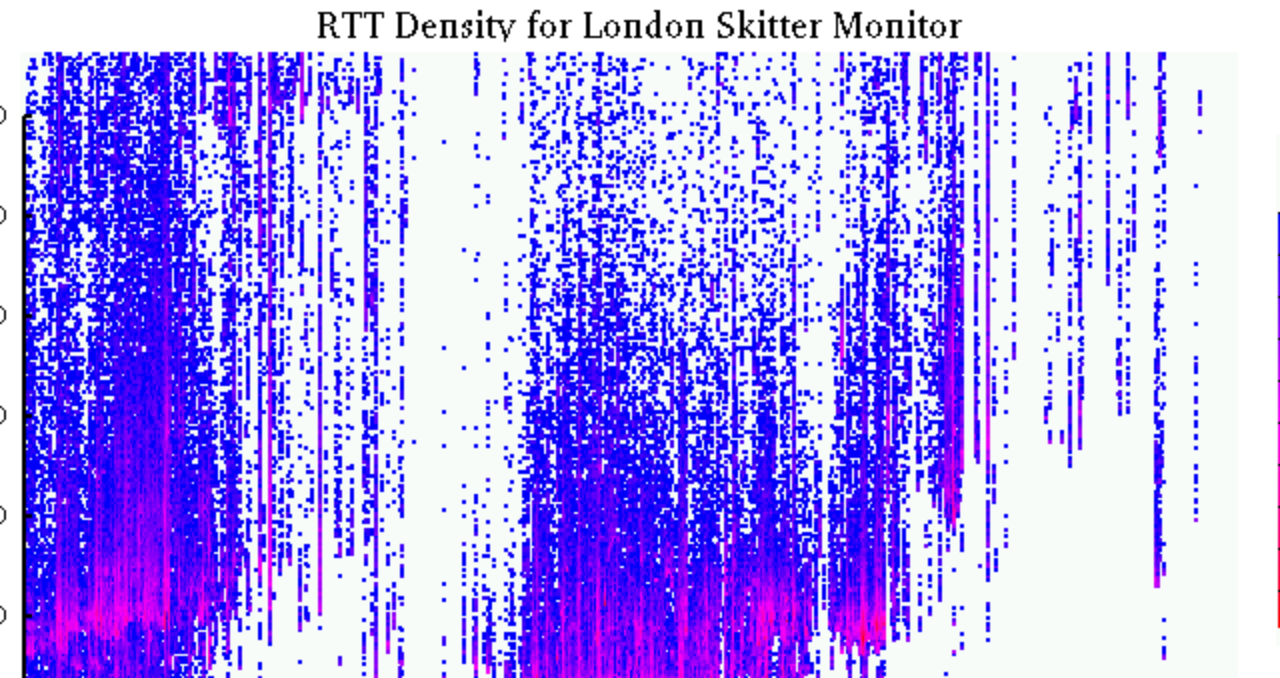
- earth circumf., X, X+Y, to-US-fr



skitter: rtt versus distance

london source

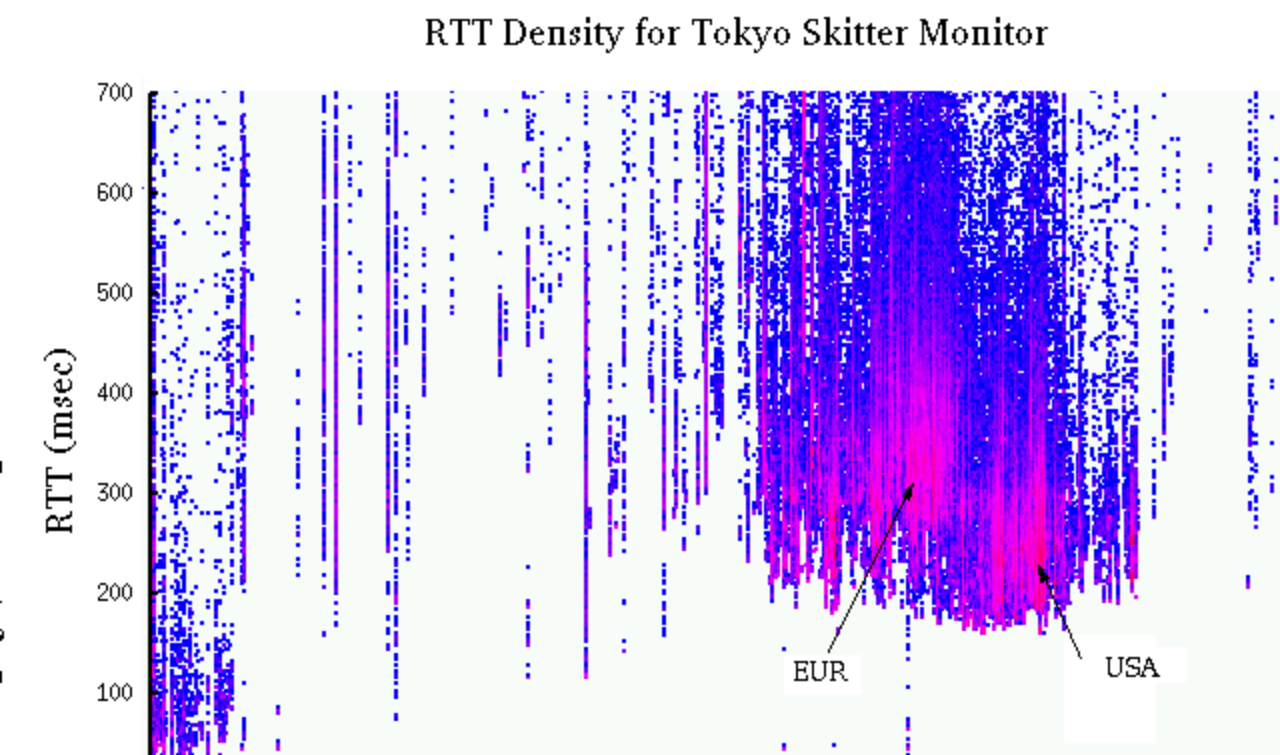
- lower band: directly connected
- upper band: thru US to rest of Europe



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

available data: (passive) header traces

coral: oc3/oc12 'real' networks

- HPC sites: http://moat.nlanr.net/Traces/

tcpdump: campus/corporate sites

- http://ita.ee.lbl.gov/html/traces.html

public tools: passive data collection

- snoop
- netramet
- tcpdump
- tcpdpriv
- tcptrace
- libpcap
- libcoral
- coralreef

public tools: passive

- snoop
  - ships with solaris
  - supports filter patterns
- netramet
  - RTFM flow meter
  - Coral interface
  - lots of supporting utilities
  - used for accounting in NZ, elsewhere
- tcpdump
  - LBL
  - ships with free Unices
  - supports filter patterns
  - understands many network types
  - well patched, supported, leveraged, evolved
  - 2 tools: capture (pcap) and display (text output)

public tools: passive -- (cont.)

- tcpdpriv
  - like tcpdump for privacy-concerned
  - encodes IP addresses/ports/strip payload, etc
  - option to exit after N packets or M seconds
  - deals with pcap/tcpdump files, same bpf filter patterns.
- tcptrace
  - post-processes tcpdump files
  - examines TCP sessions
  - measures RTT between endpoints
  - generates xplot input files (session dynamics)

public tools: passive -- (cont.)

- libpcap
  - device-independent capture library
- canonical filter rules
- framework for many tools, e.g., tcpdump

- libcoral
  - similar goals to libpcap
  - supports [ATM] cell-based networks
  - equivalent perl API

- coralreef
  - built on libcoral
  - collection of analysis tools
  - automated reports

data sets: research issues

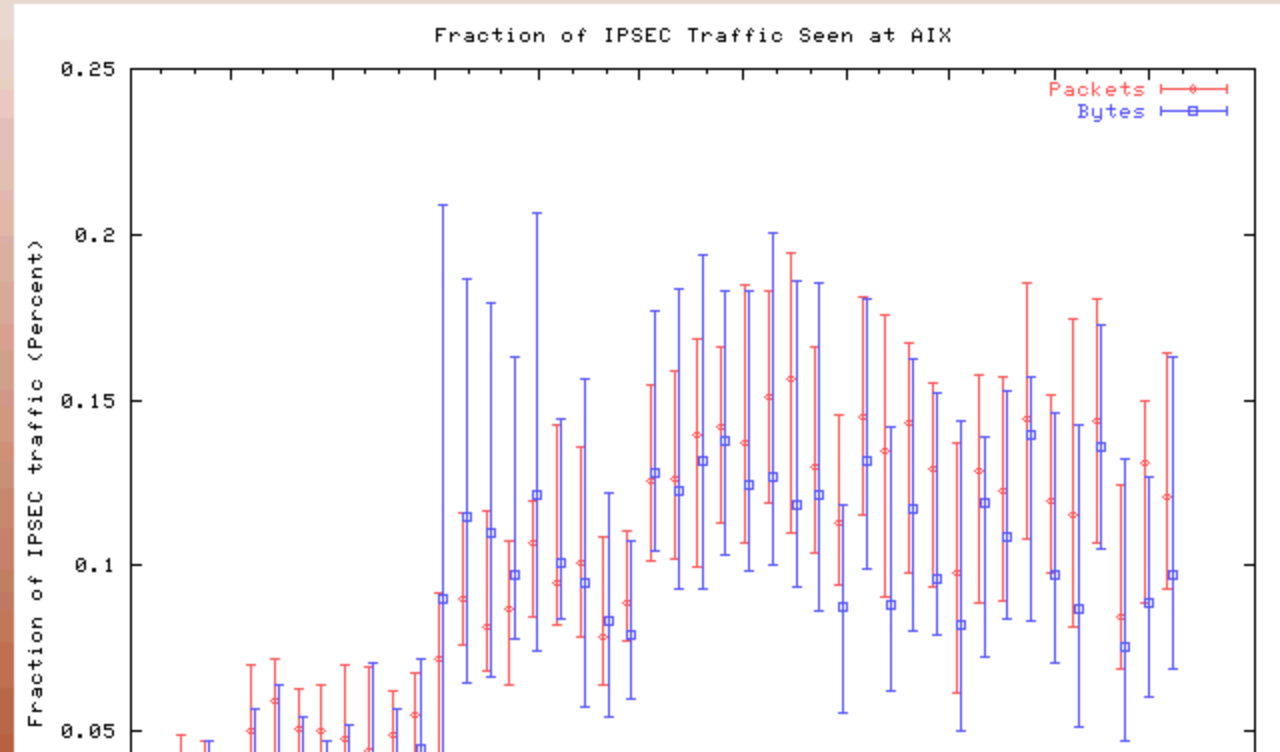
- IP address protection
- protected environment
- payload
- disk space

dataset issues

- location
- clocks
- filtering/encoding
- size
- damage

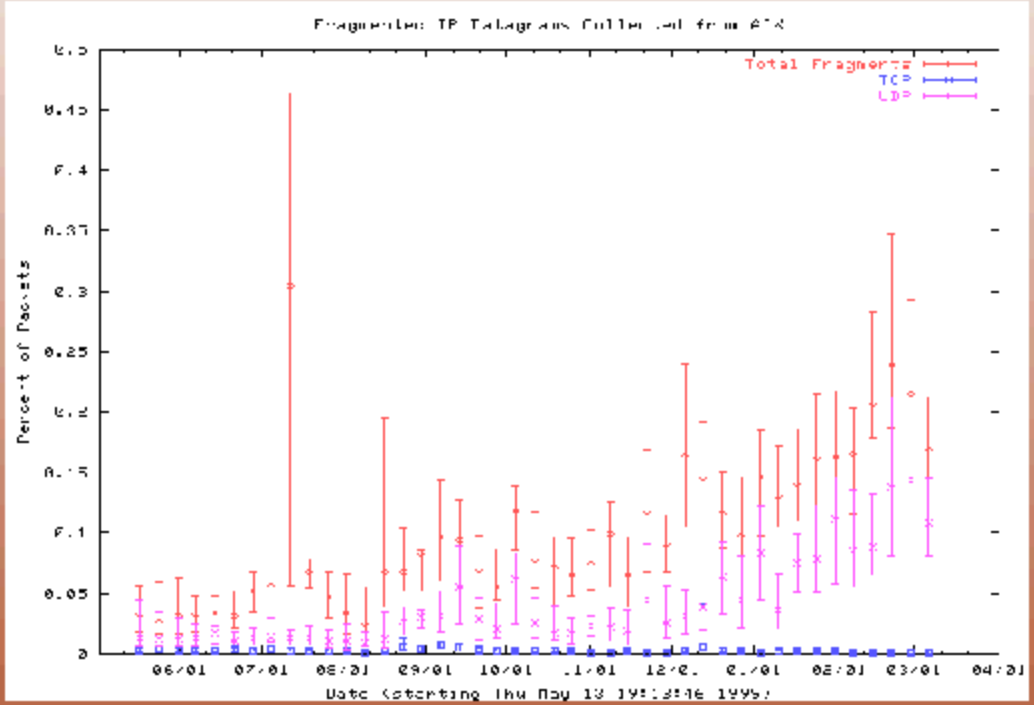
workload: AIX–MAEW ipsec (ah/esp)

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



workload: AIX–MAEW fragmentation

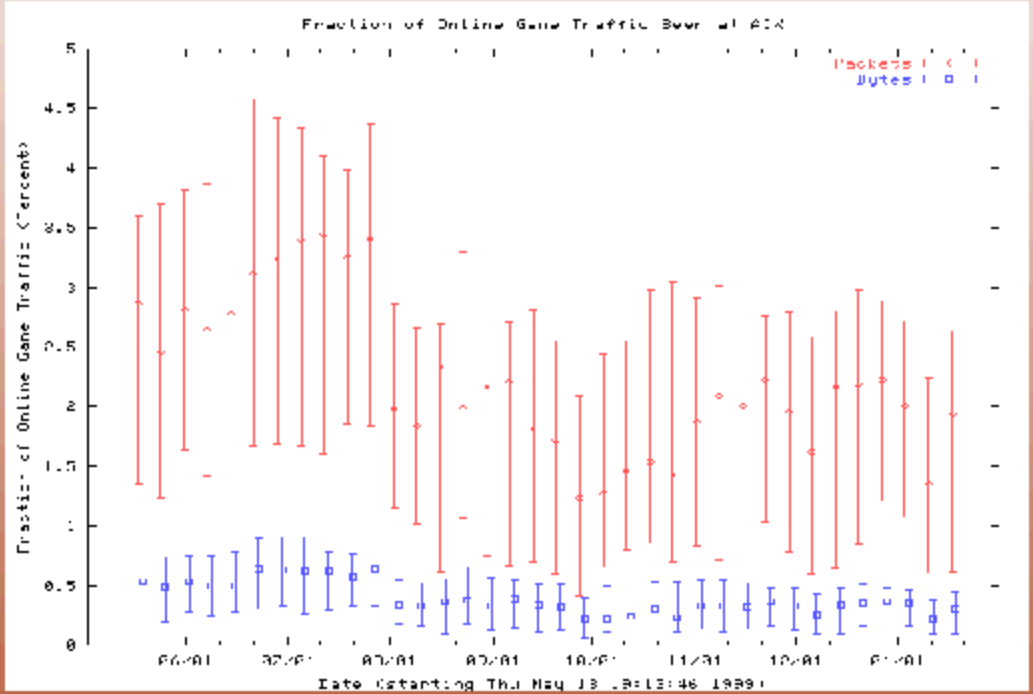
relevant to recent IP traceback techniques [Savage00]
definitely on rise (from UDP) at AIX
almost no TCP frags (MTU disc + small pkts)



workload: AIX–MAEW online game traffic

decline for games included

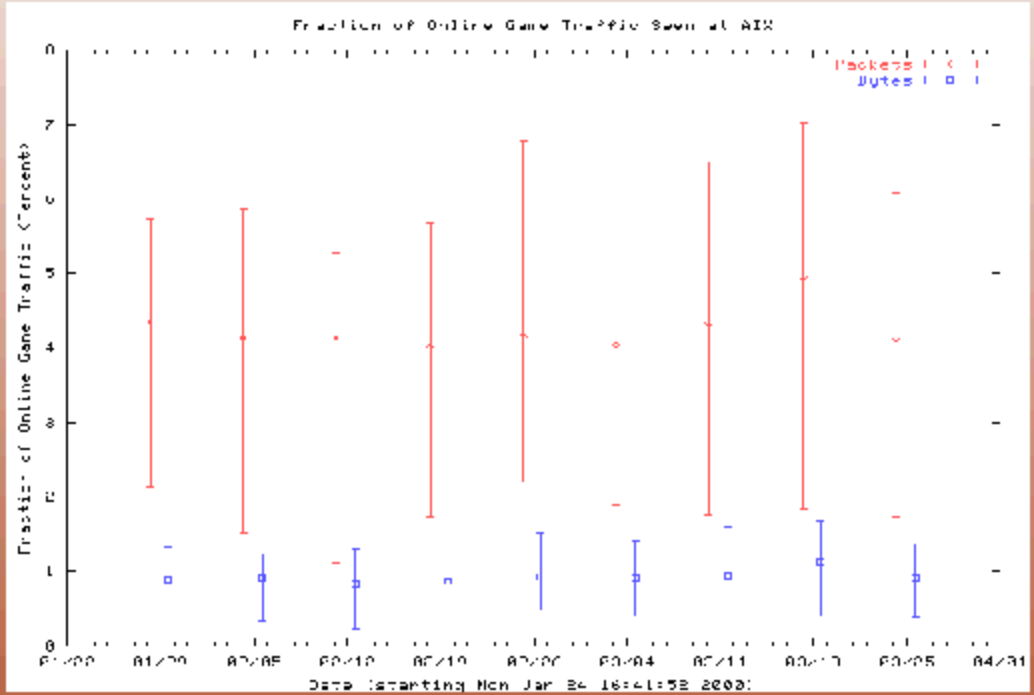
Starcraft, Quake II, and QuakeWorld (a variant of Quake II) popular last year, then other games take over?



workload: online gaming cont.

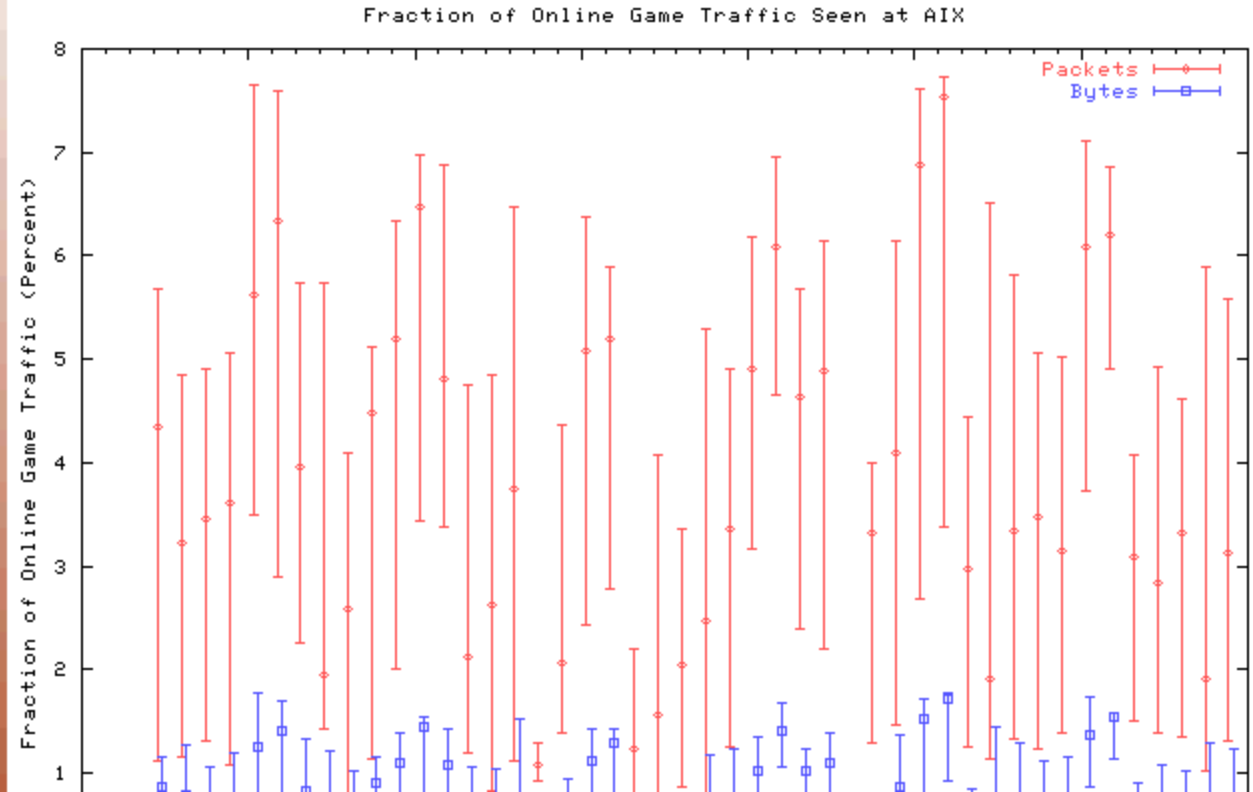
looking at new games plus old

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



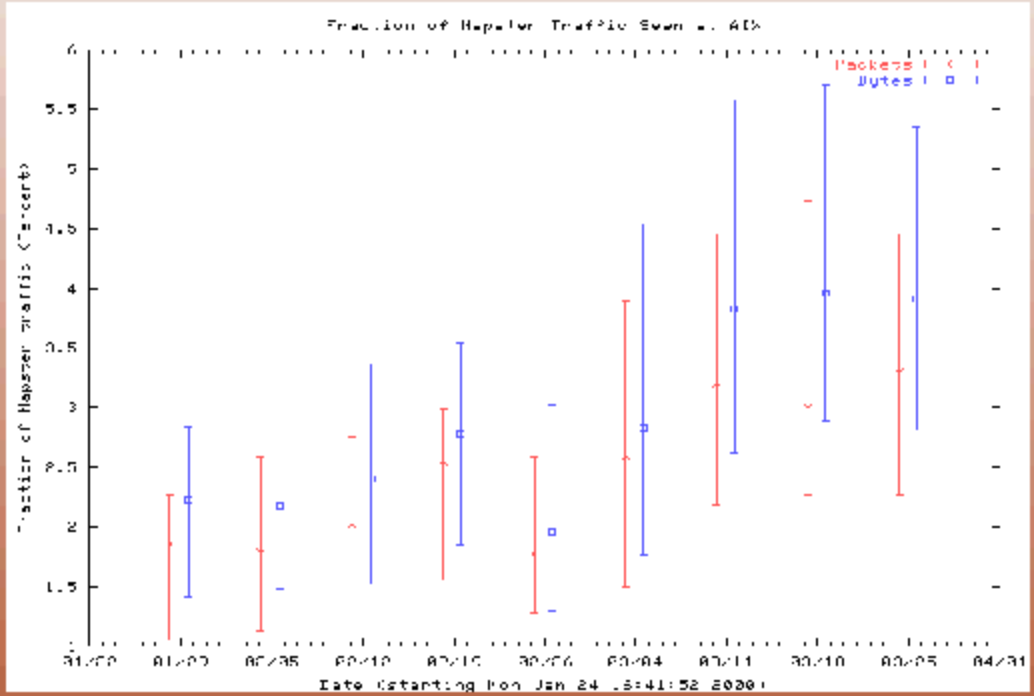
workload: AIX–MAEW gaming trends

clearly more popular on weekends (nearly double!)



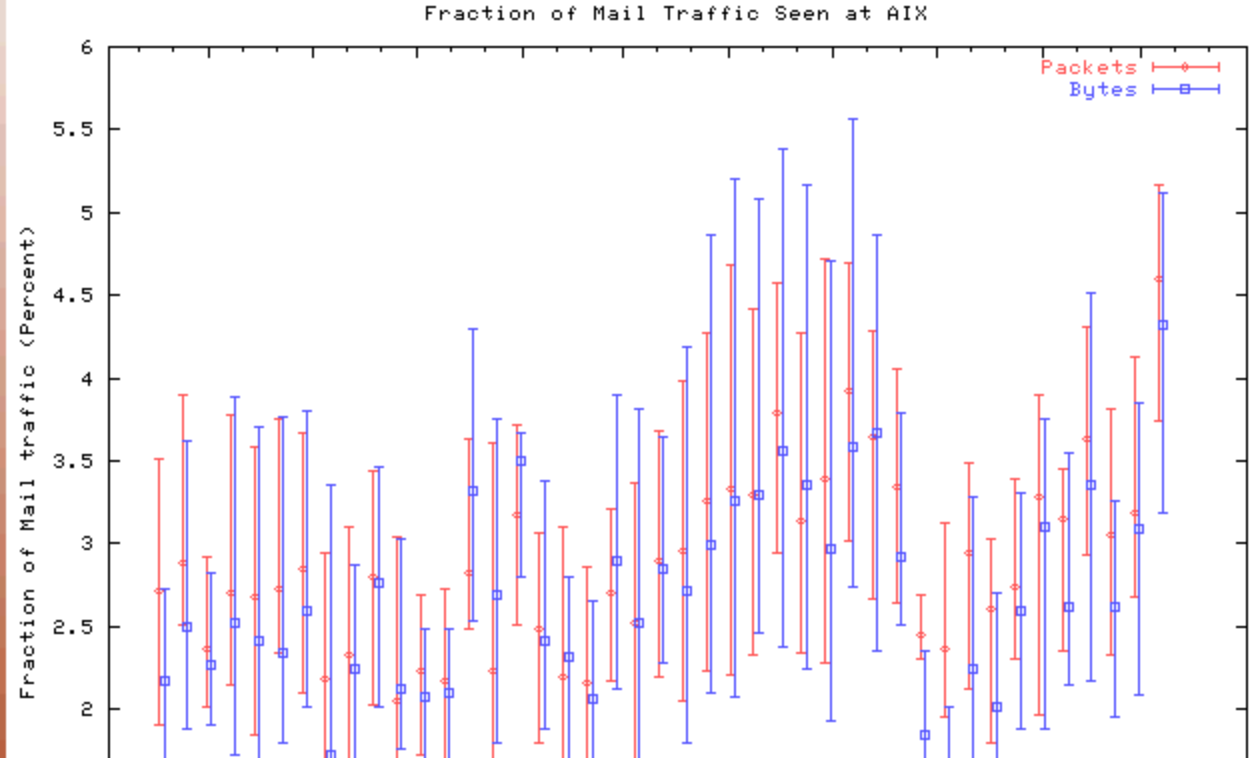
workload: AIX–MAEW napster traffic

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 AIX 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 data: 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 data: challenges

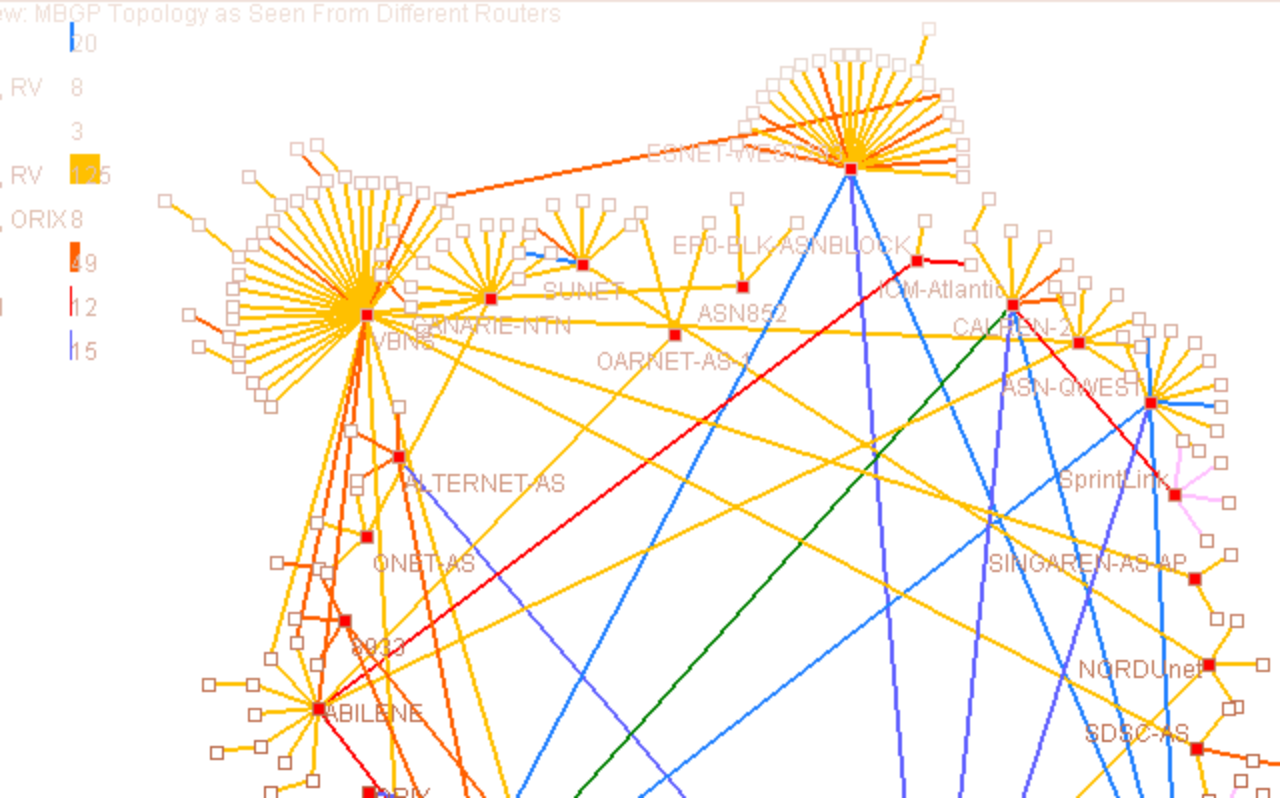
- id and present 'useful' workload metrics, particularly given persistence of fire-fighting environment
- id significant patterns, timeframes, correlations
  - vary by user need
  - change as technologies and 'net change
- methodology has many weaknesses
  - dynamic port negotiation (napster)
  - tons 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 data

- not much real–time instrumentation on routers
- UO's route–views http://www.antc.uoregon.edu/route–views/
- Merit's IPMA http://www.merit.edu/ipma/

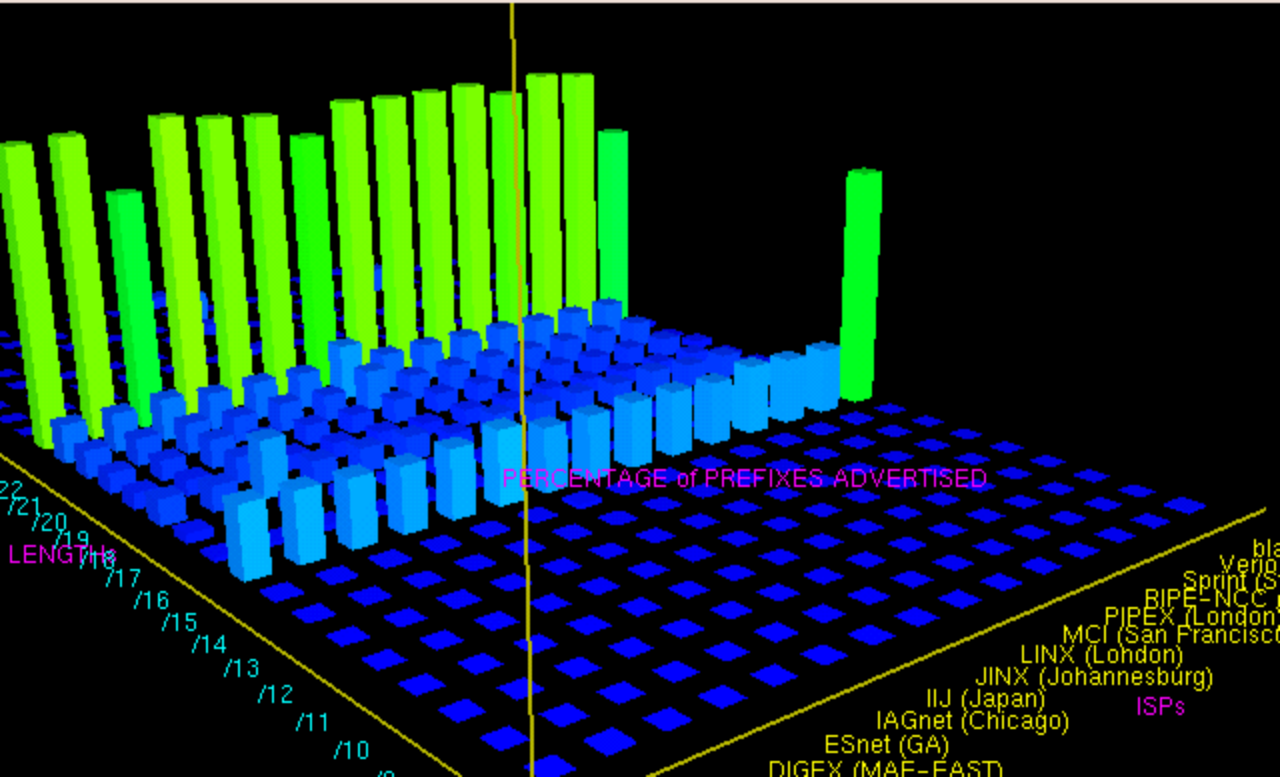
routing: differencing routing tables

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



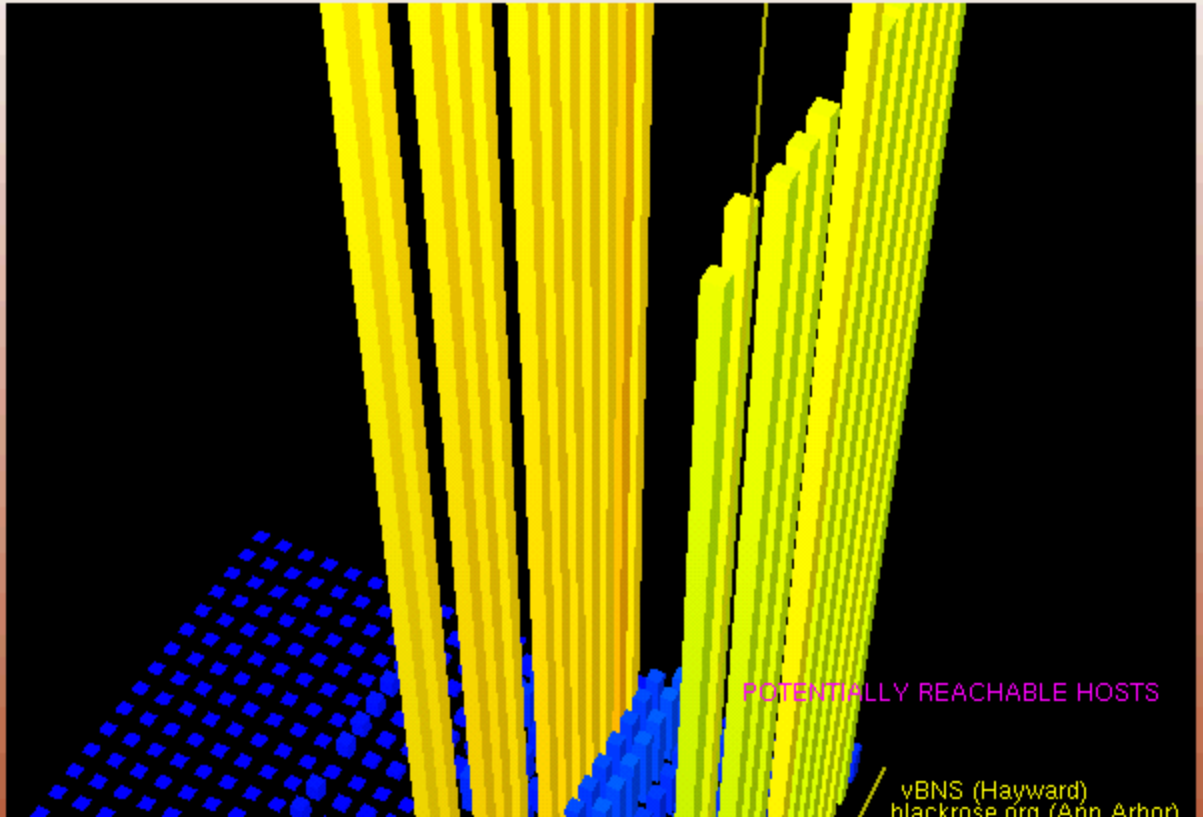
routing: address consumption

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



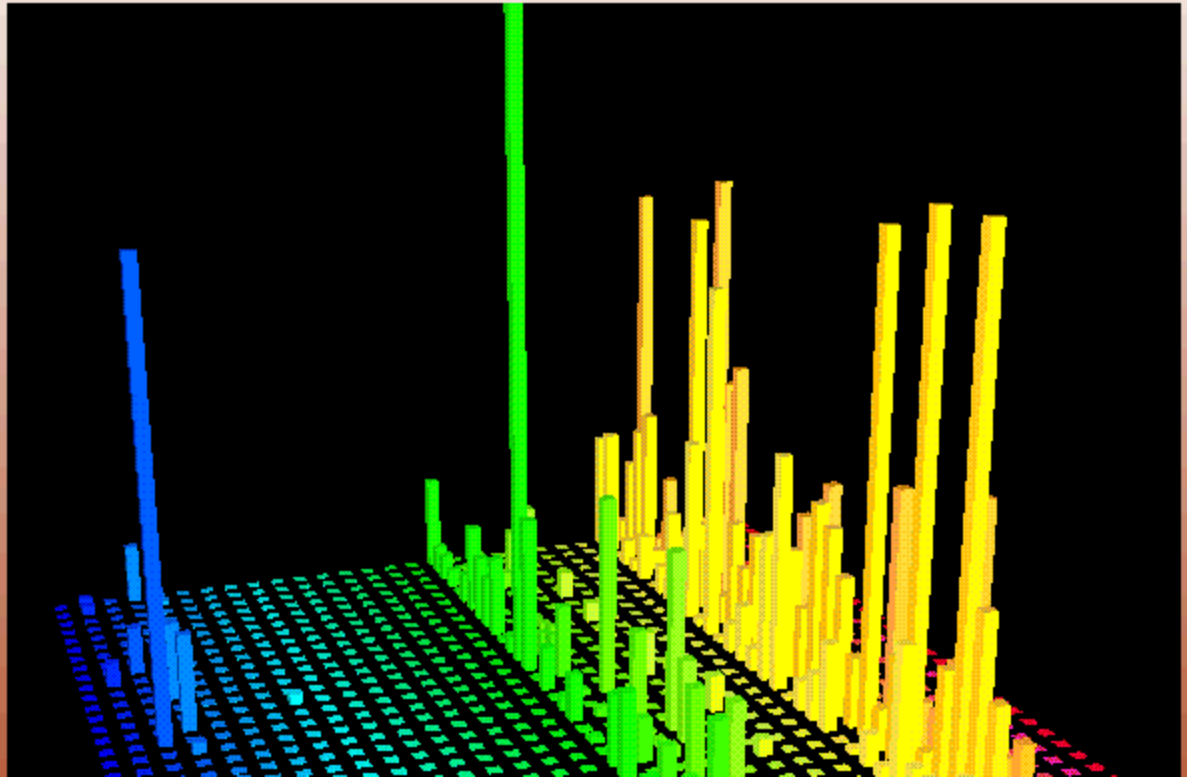
routing: address consumption (#hosts)

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



routing: address usage of *traffic* sample

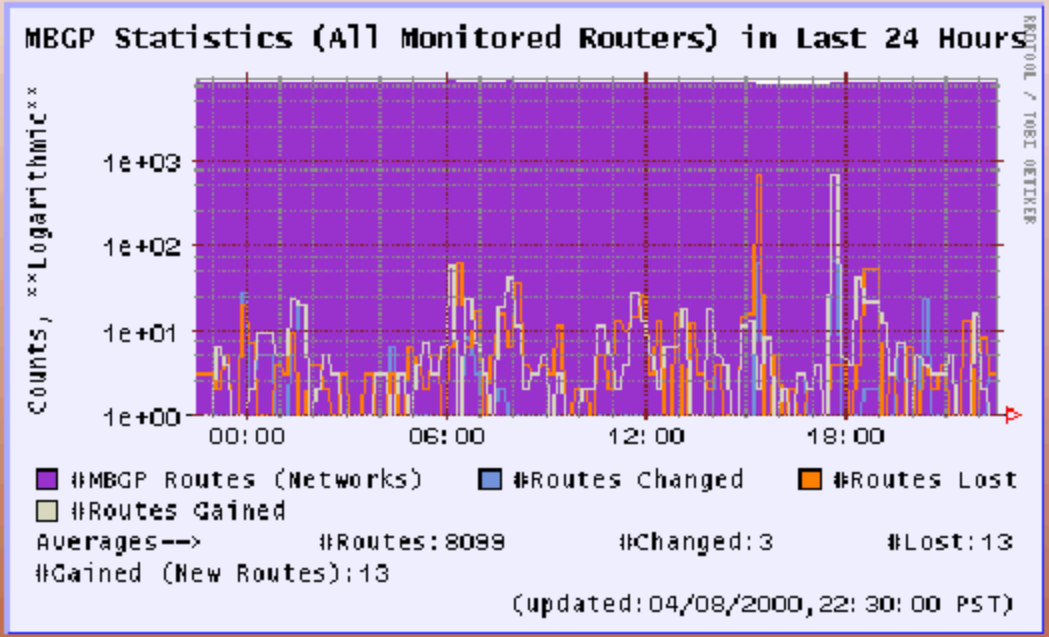
32x32 'bitmap' matrix of address space
height is % packets with src IP in that address block



multicast data sets (using mantra)

http://www.caida.org/tools/

- daily updates
- mbgp, msdp statistics, topology maps/diffs



routing: 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: 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)

setting expectations

rule 1: no magic data sets (not so far anyway)

the so–called science of poll–taking
is not a science at all
but a mere necromancy.
people are unpredictable by nature,
& though you can take a nation's pulse,
you can't be sure that the nation
hasn't just run up a flight of stairs.

--E. B. White New Yorker, Nov 1948.

www.caida.org/Presentations/

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