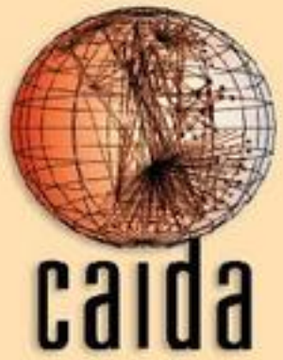


internet measurement:
myths about Internet data

it ain't the things you don't know that hurt you,
it's the things you know that ain't so.
- mark twain

july 2002
ucsd/sdsc/caida
kc@caida.org
http://www.caida.org/outreach/presentations/

what i mean by `myth'

- if you google for "Internet myths", you'll get lots of figments about Internet marketing/sociology, like
  - it's cheap to do business on the web.
  - advertising is flocking to the web in record numbers and will be its savior.
  - you can give away the merchandise as long as you generate enough eyeballs because one day you will monetize those eyeballs.
  - if you have a clever URL, they will come.
  - people will never pay for content over the web.
  - traditional advertising brings eyeballs which generates much traffic.
  - people like to shop on the web (<-- that's a good one).
  - it costs nothing to get a site up and running.
  - the web is a reliable commercial activity.
  - just you wait, profitability is right around the corner.
    - **http://www.thestreet.com/comment/wrongtactics/786636.html**
- these are not `myths' since noone actually believes them
- these are called fantasies
  - people want them to be true... (or more sustaining)
  - get return for convincing someone they're true
- myths: things people actually believe but that are wrong

fantasies vs myths

fantasies

- who believes:
  - marketing, advertising people, lawyers, consultant (consenting) adults
- who gets hurt:
  - marketing, advertising people? (no comment..)

myths

- who believes:
  - researchers, vendors, policymakers, journalists, secretary of defense
  - potentially: marketing, advertising people, lawyers, consultant (consenting) adults
- who gets hurt:
  - packets (dropped)
  - engineers (paged)
  - protocol developers (in worst case they invent stuff like atm, mpls)
  - grad students (useless dissertations, sub-employability, lost decades of youth)
  - economy (irrational speculation in capital markets -> global recession)

Internet myths relevant to engineering

- workload: (besides basic traffic growth fiction)
  - level and nature of fragmented traffic
  - increase in flows as bandwidth grows
  - mice vs elephants
  - private addresses in core
  - prevalence of encrypted passwords
  - applications can be identified (much less controlled)
  - multicast traffic, flows, addressing
- performance:
  - DoS attacks affect only large sites
  - geography not correlated with latency
  - DNS system performs well
  - single router can't trash the Internet
- topology:
  - Internet topologies, object sizes follow power laws
- routing:
  - routing tables reflect Internet topology
  - intra-country traffic stays there
  - AS path length is decreasing
  - small providers and multi-homing (more specifics) cause all the churn

why so many? no real measurement. (so no real surprise...)

Internet's resistance to modeling/measurement

evolution-based (good!) reasons

- protocols, technologies, applications, users
  - independently developed and deployed/connected
  - by no means synergistic
  - by all accounts rapid
  - 'punctuated' but no equilibrium
  - "have done fine without modeling so far" (let's wait till modeling is cheaper than bandwidth)

but simulation/analysis validation (& lately engineering/billing/[homeland.]security) needs data

- right granularities hard to come by
- measurement technology just not there
- argument for it also not there
- "helps everyone", but who pays?
- losing battle?

measurement tools lack (related obstacles)

- 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 repair
- communication of useful results

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/accomplish measurement?
(when market forces are detrimentally torqued)

alternatives:

- 1) tools that positively affect an ISP's balance sheet
- 2) regulatory intervention

what happened instead of measurement

a. odlyzko's "myth of Internet growth" study (nov 2000)

- plus great assessment (...) of larry roberts 2001 caspian.goo
  - (see also his recent lightreading/other statements)
- 'traffic doubling every 90 days'
  - maybe for a few months in 1995-1996
  - in reality, no real data since 1995 (nsfnet sunset)
  - more like every 12-18 months for rest of 1990s
- financial markets (at least in US) believed (bubbly!) estimates
- over 6 years, that would have meant a factor of 16 million
  - assume (generously) 500M users, 1.5Mbps per user around the clock
  - and yet we're mostly still using 28k modems, & only for an hr/day, & ave 5k bits/sec
- come again? the math just does not work out
- it took 5 years for true traffic growth data to finally manifest itself
  - (since providers would not release data, if they even had it) via other metrics (hardware and bandwidth sales) required in annual reports to SEC (closest we have to an Internet Measurement Commission)

that's actually an embarrassingly pathetic willingness
to ignore real data (or just invent it)

living in a mythical world: tradeoffs

costs

- tech stock bubble? (hey infinite demand is infinite jnpr stock price)
  - really takes new technologies a decade to penetrate
  - web was exception (when it was young/free), Internet is not
- retarded technical developments
- negligence of what users what, and likely to get
  - community gets mired in sub-necessary QOS hubbub, ATM, GMPLS

benefits

- unparalled platform for innovation
- open standards, rapid development of new services
- big empty pipes was key factor in supporting [r]evolution
  - pipes wouldn't be empty for grad students (napster,kazaa) if the myths had been true

lessons

- 25 year contracts for pipes should be amortized over 3
- come to terms with a much looser definition of `capacity planning'
- simplify engineering (atm/sonet --> IP over WDM, GigE.)

(first commandment: Thou Shalt Get Rid of Layer Goo)

living on borrowed time in a mythical world

(opportunity costs of not measuring)

- three `waves' of Internet applications/usage
  - first wave: shared (remote) use of computers
    - **telnet, email, ftp**
  - second wave: client/server model, formatted languages
    - **web**
  - third wave: collaborative, peer-to-peer, interactive
    - **napster, imesh, kazaa, gaming, video**

emergence of third wave (`ngi') will require more real-time interaction with and reaction to network status

- growth of these applications will be self-limiting (by user frustration with performance) unless we have either:
  - a better grip on measurement
    - **either done by the applications themselves (e.g., vat)**
    - **or via some other middleware aspect of the infrastructure**
  - or no service-affecting queueing anywhere in the network
    - **seems unlikely, even with lots of empty pipes**

four areas of measurement (and thus myths)

- workload characterization (passive)
- topology (mapping, path dynamics)
- performance evaluation (active, passive)
- routing (dynamics)

caida focuses on

- measurement tools (prototypes)
- macroscopic (or macroscopically relevant) analyses
- identifying priorities and obstacles

workload measurement: in the Internet `core'

(background only)

- caida/waikato coral+dag oc48mon
  - first and only OC48 flow monitor worldwide
  - caida's public software analyzes data without modification
- software implemented
  - CoralReef, NeTraMet, custom routines (CAIDA)
  - other custom/enhanced routines by U. of Waikato, others
  - darpa/nsf/caida members funded
- software, data analysis, viz tools all prototypes
  - commercial spinoff for the cards (www.endace.co.nz)
- but btw backbone core now needs oc192/oc768 monitoring
  - there's not yet even such thing as a 24 hour trace for an oc48 link

workload measurement: in the Internet `core'

(background only: dag oc48 capture card)

- current oc48mon system (prototype at MFN in SJC, subc/collab. w U.Waikato)
  - captures 1M packets/sec to disk (40% util. link)
  - provides highly accurate timestamping
  - .5Mp, 1Gbps (125MB/sec) each direction
  - avg pkt size 370, 590 bytes (210k, 340k pkts/sec)
  - 64 bytes/record -> 6-9X compression over link load
  - problems: bursts of small packets cause machine thrash
  - http://dag.cs.waikato.ac.nz/
- upgrading oc48mon this qtr to house (bigger) Dag4.10 cards
  - dual-Pentium (Intel) processor on tyan S2510
  - 1GB of RAM
  - floppy, cdrom
  - IDE/ATA disk drive (40GB min)
  - 6 SCSI Ultra/160 disks, 3/each SCSI channel each 18GB min
  - 4U rack mountable chassis

this will get us One Hour (and just barely, and ~50GB)
(MFN SJC 76 min @20:00 PDT 5 aug 2001 ==> 32GB)

workload measurement: in the Internet `core'

(background only: workload data analysis)

- use coralreef software suite
  - http://www.caida.org/tools/measurement/coralreef/
- obtain quantitative parameters of captured traffic:
  - byte rates and packet rates
  - flow = {src ip, src port, dest ip, dest port, protocol}
- use netgeo tool to map src/dst ip addresses to ASes or countries
  - http://www.caida.org/tools/utilities/netgeo/
- consider various aggregations of traffic:
  - applications
  - ASes
  - countries

workload myth: mice vs elephants

myth:

10% of flows cause 90% of traffic on a link (or pick x%,y%)

- data:
  - sometimes true for bytes
    - **especially if the link has KaZaa-type stuff**
  - never true for packets
    - **in any traces we've studied**
  - actual proportion of traffic (bytes or packets) covered by 90% of streams can change rapidly
    - **following changes in the applications/protocol mix**
    - **obviously depends on site as well**
    - **but also per site changes over course of day, week, month**

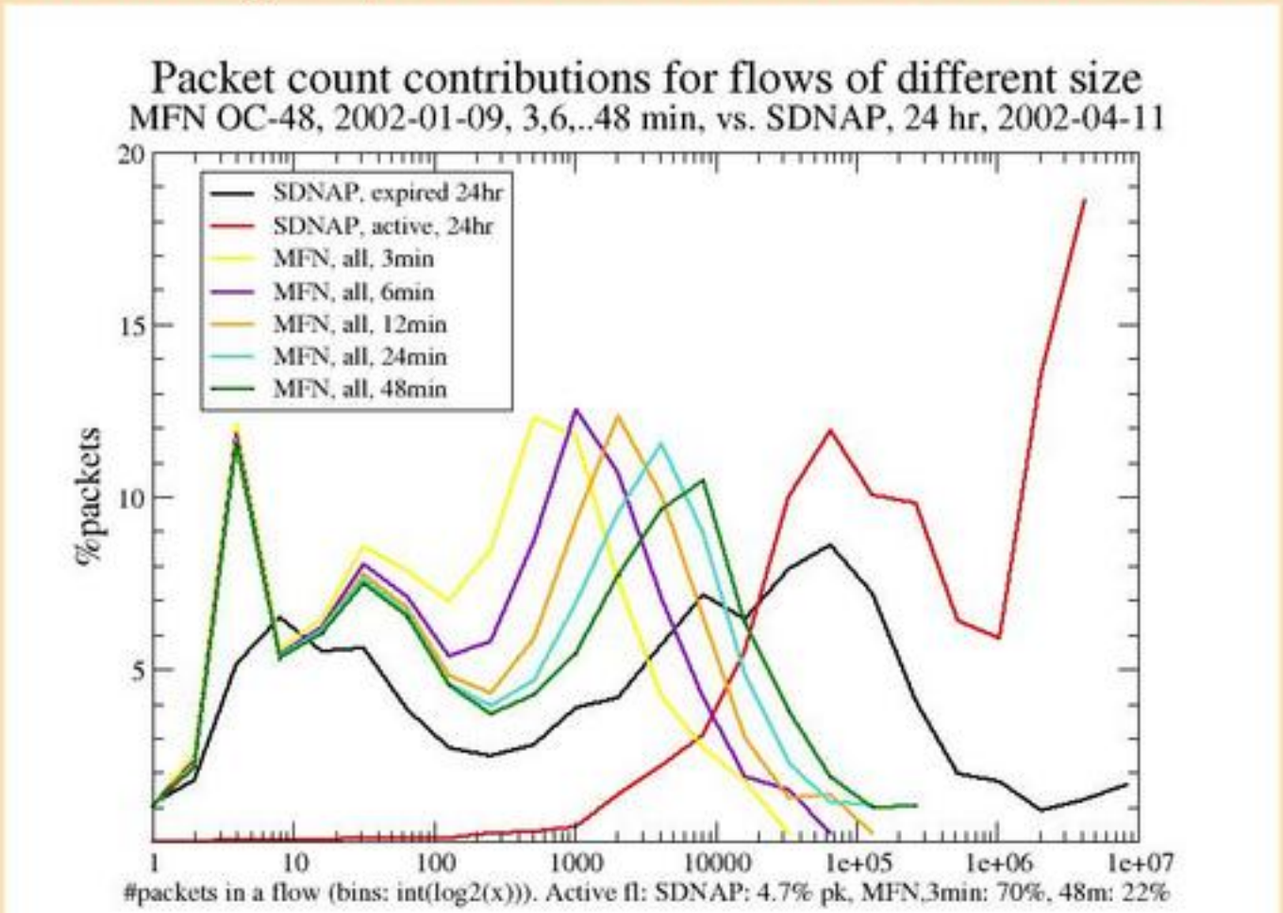
need to measure proportions before making assumptions need longer traces!

workload: mice vs elephants

longer traces are essential

- two modes of Internet usage (interactive, downloads)
  - **boundary between modes is ~300 packets (0.5 MBytes)**
- most flows on the left (by far), most packets on the right (by far)
- for a 24 hour (sd) flow table, 4.7% packets are in still-active flows
  - **50% packets are in flows with > 8192 pkts; max. flow: 9M pkts. max. active flow: 5M packets.**

upshot: do not analyze flow sizes with less than 24 hours of data

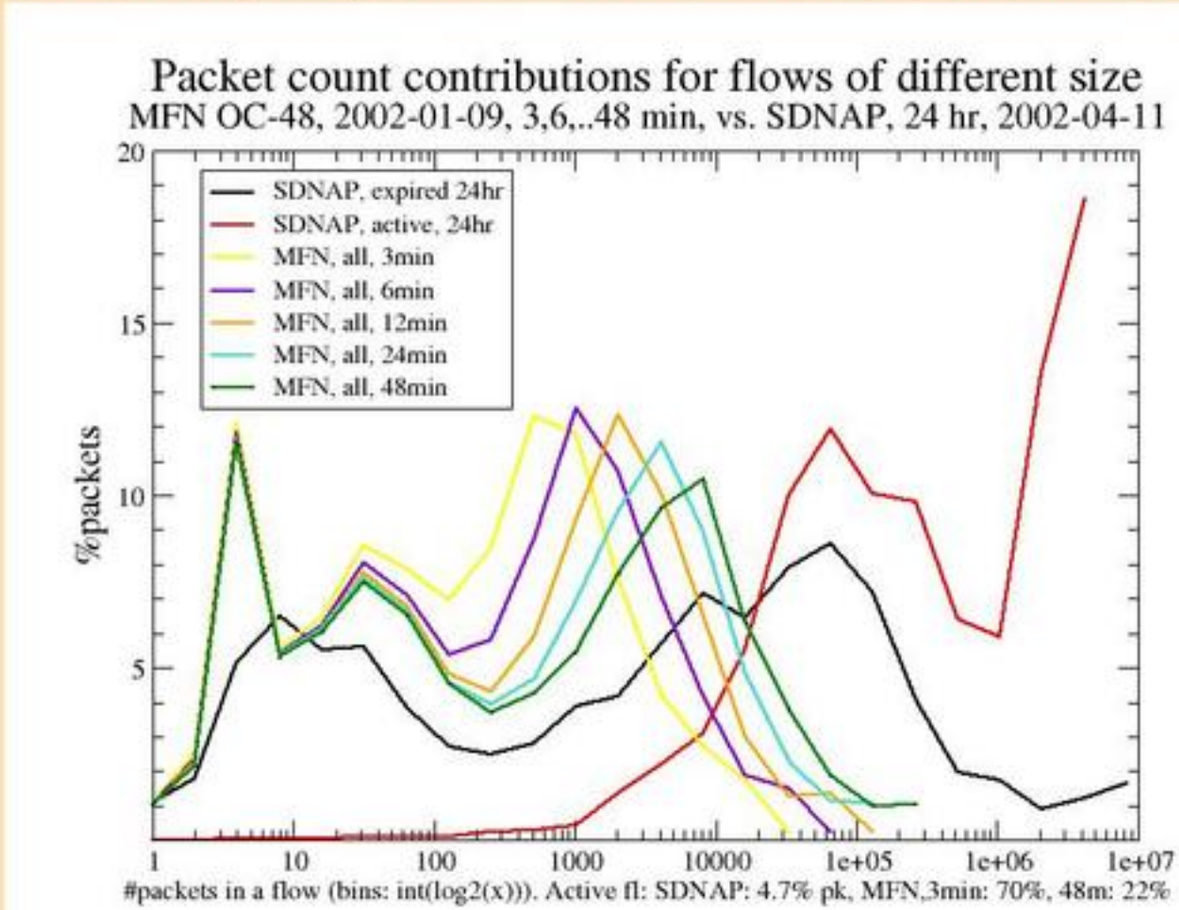


workload: mice vs elephants

longer traces are essential

- for a 48 min (sjc) packet trace (15GB), 22% packets are in still-active flows
  - **50% packets are in flows with > 1024 pkts**
- for 3 min (sjc) trace, 70% packets in still-active flows
- for each 2X in sample duration, 2X in max of pkt/flow
- convergence nowhere in sight

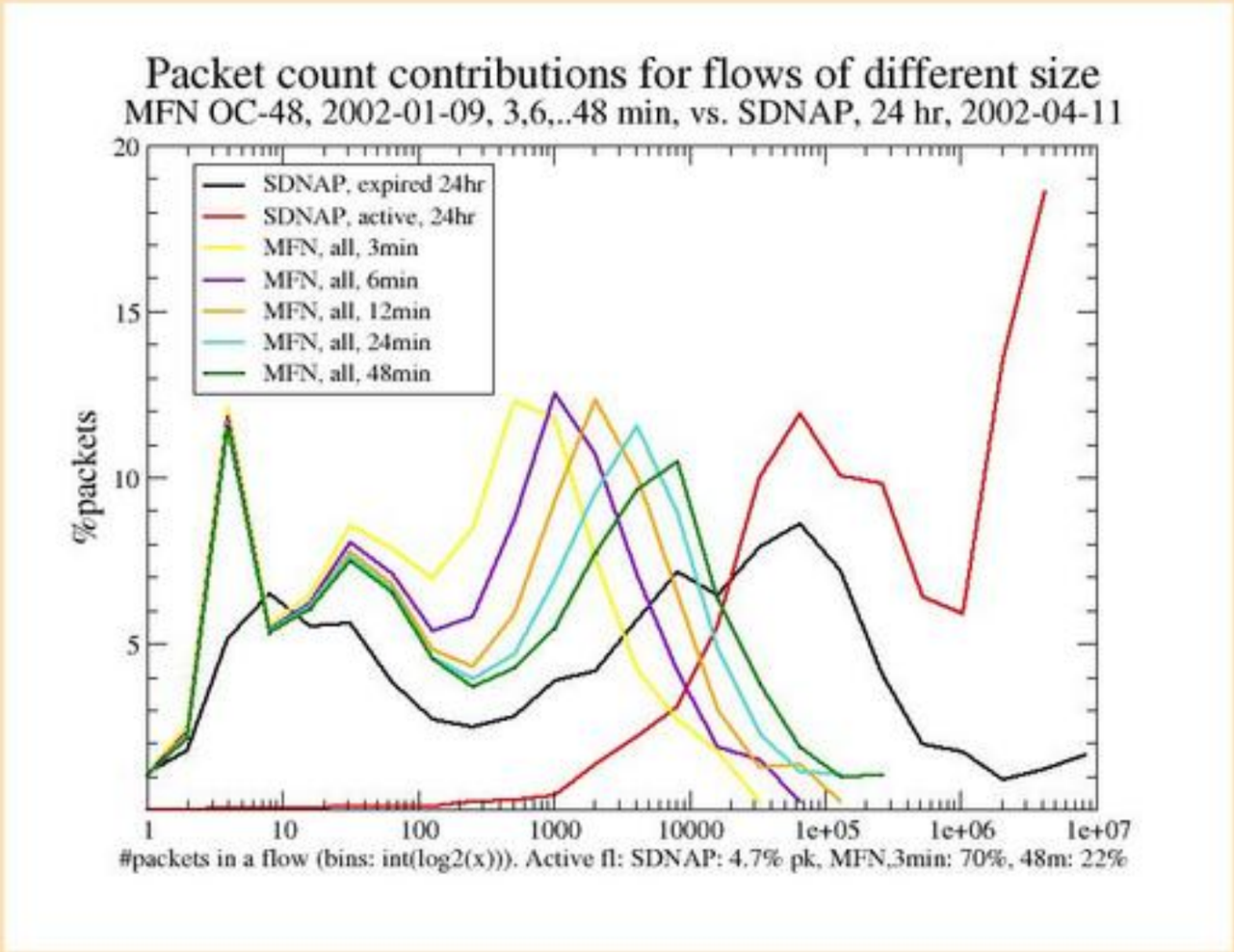
repeat do not analyze flow sizes with less than 24 hours of data



workload: mice vs elephants

(generally, we do not yet know what we're talking about)

- but we know not to analyze Internet flow sizes with less than 24 hours of data
- btw, nobody has 24 hours worth of useful core packet trace data (we're $nMs away)



workload myths: prevalence of IP fragmentation

myth:

there is no fragmented traffic

- data:
  - while it is true that overall only a small amount (1-2%) of Internet traffic is fragmented, fragmented traffic can surge to 8% by packets and 10% by bytes in some hour-long periods.
  - Some protocols, for example IGMP, have fragmented traffic far exceeding non-fragmented traffic.

myth:

fragmented traffic exists only on LANs

- data:
  - we've monitored aggregated exchange points and backbone links.

workload myths: prevalence of IP fragmentation

myth:

tcp traffic is never fragmented

- data:
  - while tcp traffic is fragmented much less frequently than other protocols due to path MTU discovery, we saw 0.009% by packets (0.019% by bytes) of fragmented tcp traffic.
  - and a majority of fragmented tunneled traffic is tcp!

myth:

nfs causes all (or almost all) fragmented traffic

- data:
  - microsoft's media player was single largest cause (52%) of measured fragmented traffic. Tunneled traffic, also a major cause, accounts for 16% of all fragment series. nfs caused 0.1% of all monitored wide-area fragmented traffic.
  - tunneled traffic (ipencap, ipip, gre, udp l2tp), icmp, and realmedia all cause more fragmented traffic than nfs (0.1%).

workload myth: private (rfc1918) addresses

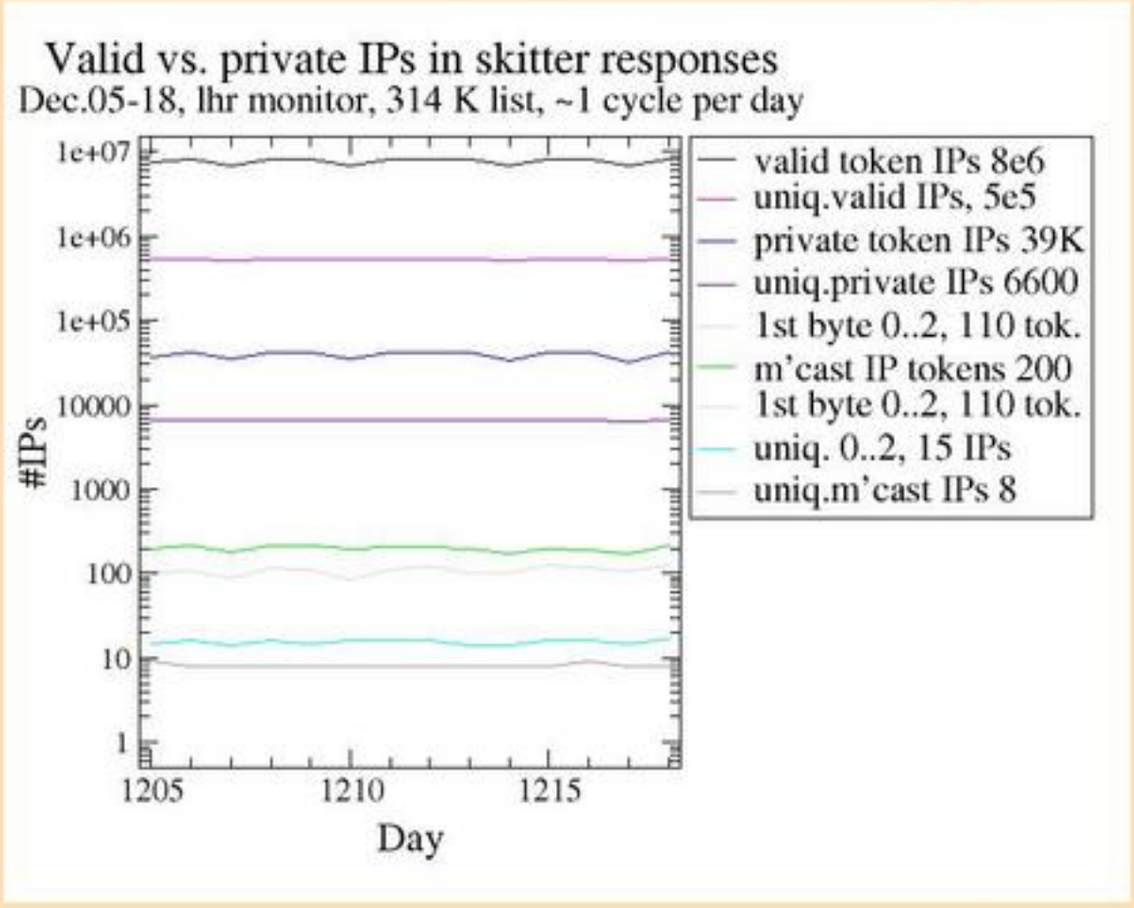
myth:

private addresses do not appear in the core

- data:
  - private addresses appear all over the place including (consistently) in queries to root name servers as do multicast and other `shouldn't be seen' junk

Broido's 1st Law:

`what should not be seen in the Internet will appear 1% of the time'

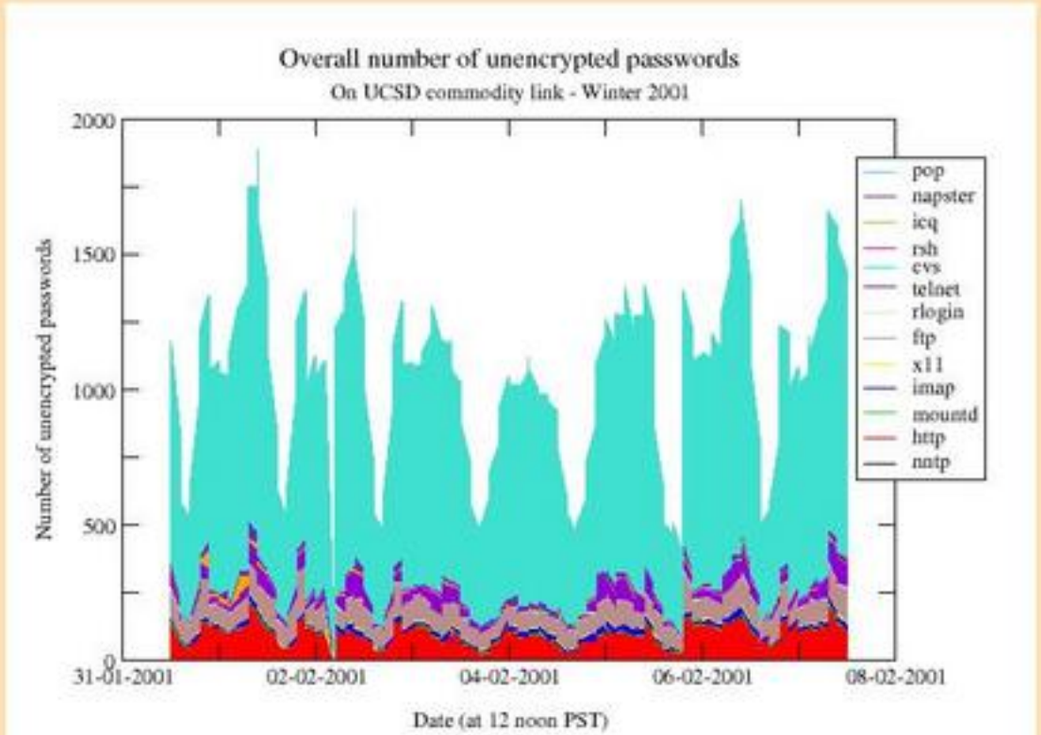


workload: prevalence of encrypted passwords

myth:

unencrypted (cleartext) passwords are mostly gone

- data:
  - most unencrypted passwords are from one source: POP
  - why aren't folks using APOP? (authentication already provided)
  - mere existence of an encryption technology is no guarantee of its adoption



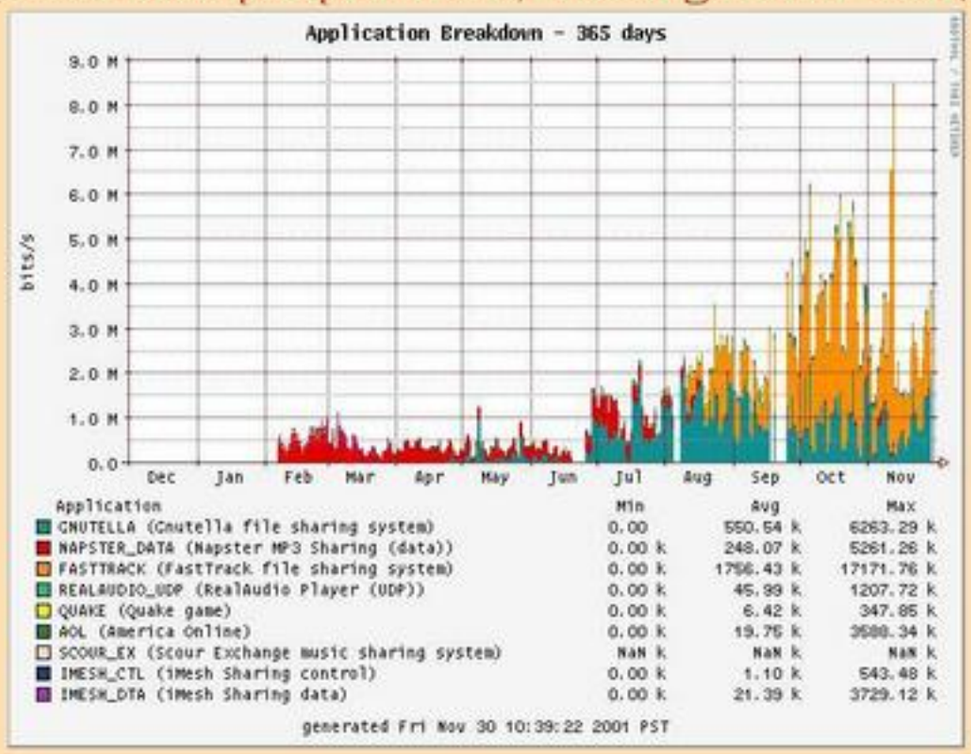
workload myths: p2p file sharing

myth:

the US government can stop file sharing
(or `legislation takes less time to write than code [v.t.]')

- data:
  - napster: red
  - kazaa (fasttrack): orange (post-federal.judge napster.shutdown decision)
  - admit it's in the fantasy category, myth might be `currently no killer app'

...in an expanding system, such as a growing organism, freedom to change the pattern of performance is one of the intrinsic properties of the organism itself...



workload myths: killer apps

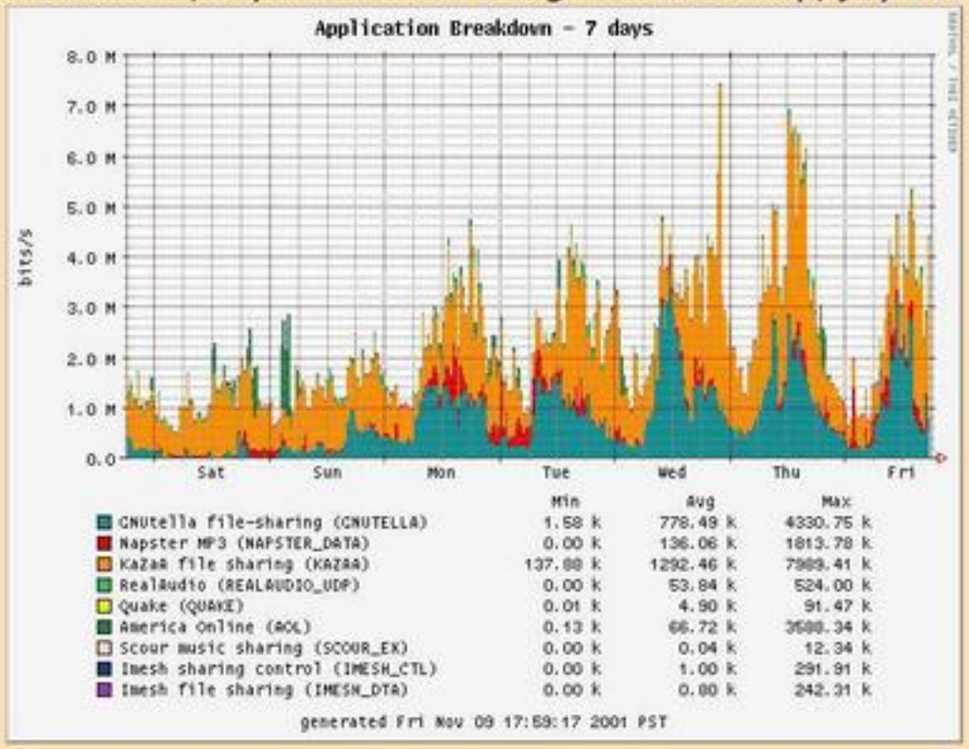
myth:

we're in between killer apps

- data

how do you know when something is a `killer app'?
when every university tries to stop it and _can't_. _that's_ how you know
it's a killer app. that it takes a federal judge to threaten to put you in jail
if you don't stop. _that_'s how you know it's a killer app!

-eric schmidt, keynote for dns navigation workshop, july 2001

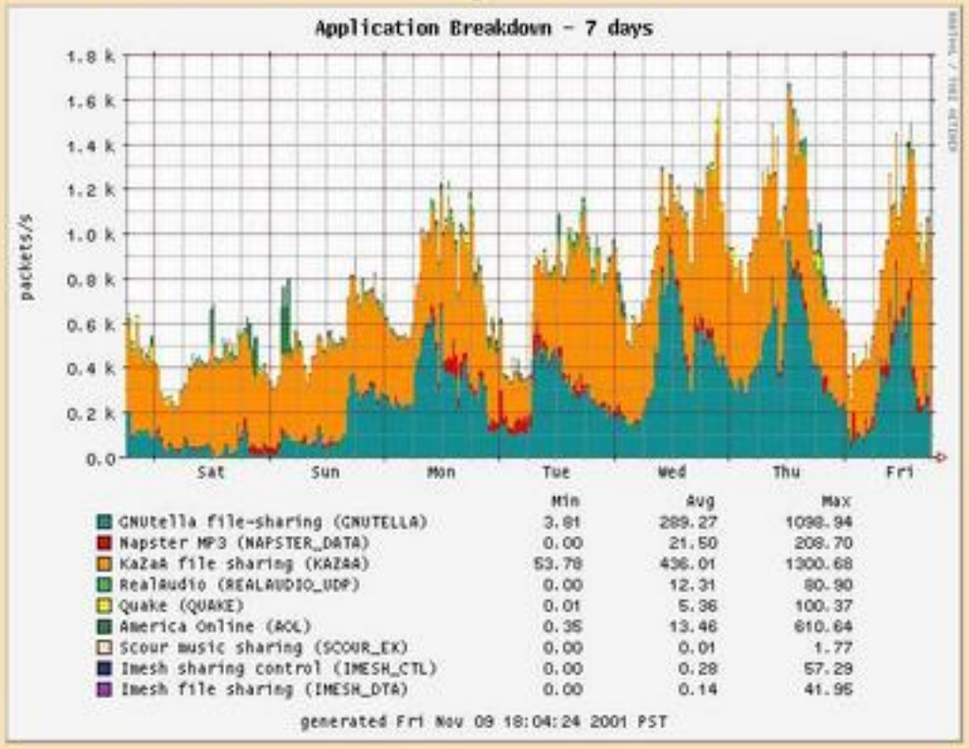


workload myths: code vs law

myth:

we can handle the few bad seeds with legislation

- data
  - not just a few huge packets sneaking in
  - note similarity to gopher/web transition (patent/port# control)
  - (not that anyone would know via measurement... ask Internet historian)
  - code seems to get written faster than legislation

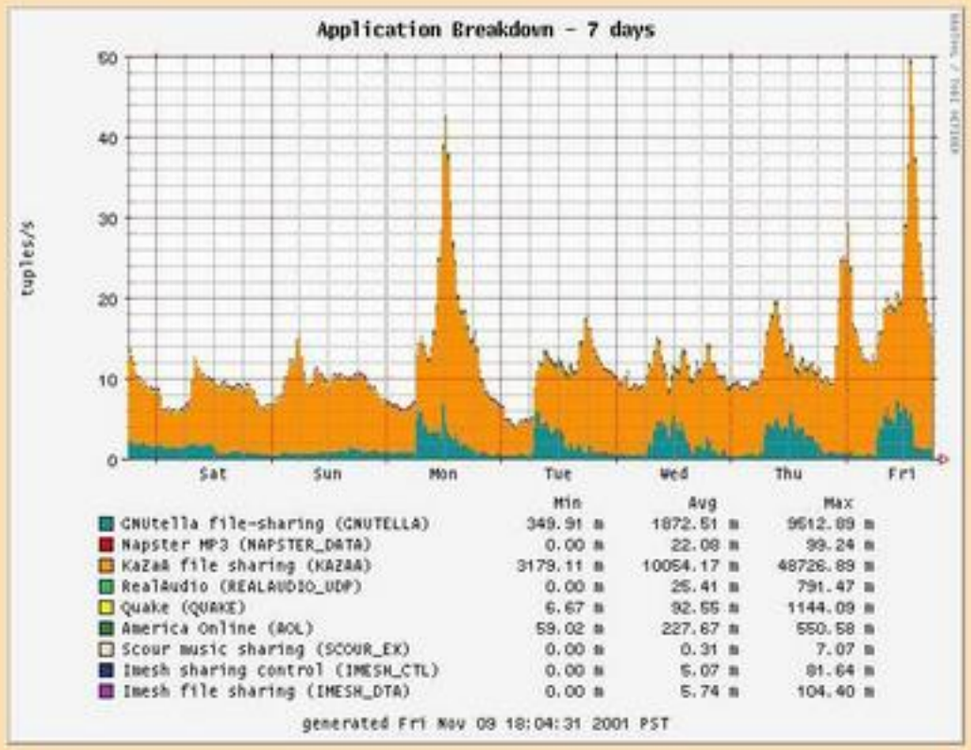


workload myths: code vs law

myth:

govt can stop file sharing/no killerapp

- data
  - not just a few punk users
  - compare how different apps affect network... especially bytes vs. tuples.
  - gnutella/fasttrack: both big flows; fasttrack (kazaa): lot more connections
  - uh, `no killer app'?
    - **(what, still waiting for multicast?)**



workload myth: multicast traffic

(background only)

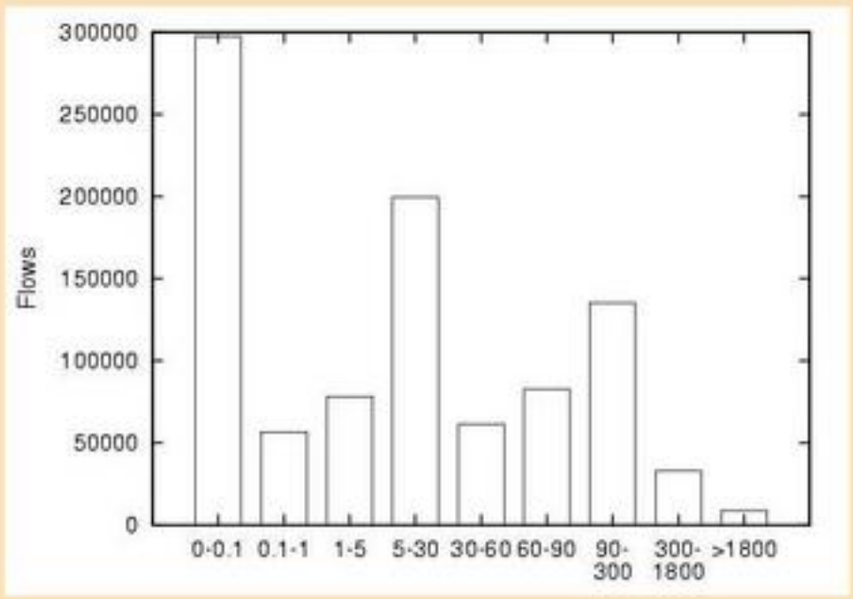
- month-long multicast flow monitoring project
- passively monitor OC-12 links connecting backbone to peer multicast networks and to customer aggregation switches
- 4M flows, 12G packets, 11T bytes

workload myth: multicast flows

myth

multicast flows are long-lived

- data:
  - 75% of flows were a single packet
  - majority of flows short-lived
  - filter out single packet flows and multicast control packets to arrive at "representative" multicast traffic
  - filtering preserves 99% of the packet and byte counts but only 14% of the flows



workload myth: multicast flows

myth

multicast flows consist of small packets and are never fragmented

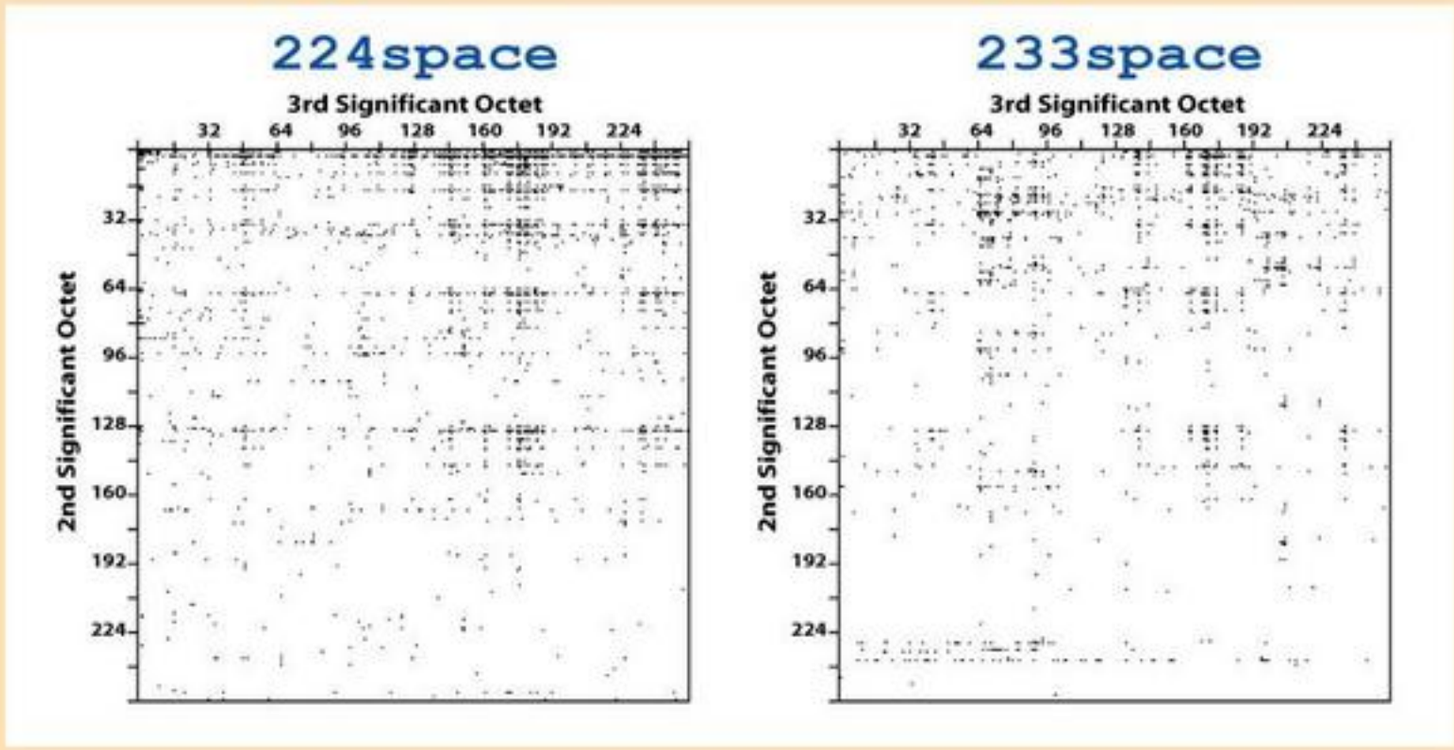
- data:
  - strong packet size modes at 1480 bytes (presumably to avoid fragmentation)
  - 0.5% of the packets were fragmented
  - 3% of the packets had the 'don't fragment' flag

workload myth: multicast addressing

myth

multicast addressing is efficiently allocated and assigned

- data:
  - psychological disposition to use beginning of each /8 multicast block
  - reserved IANA blocks indiscriminately used
  - improper use of GLOP and EGLOP space

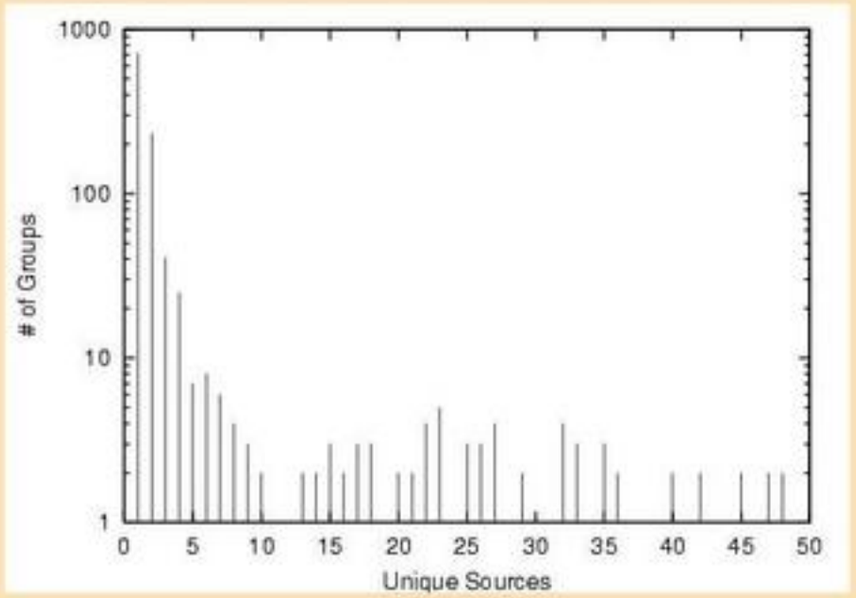


workload myth: multicast technology

myth

multiple source multicast is popular and architecturally fundamental

- data:
  - only 1% of the groups observed ever had multiple simultaneous sources
  - largest number of simultaneous sources was 33 (Access Grid)



performance myth: DoS attacks

myth:

flooding DoS attacks (randomly spoof src, main attack type)
only affect large commercial sites, are long in duration
and at extremely high rates

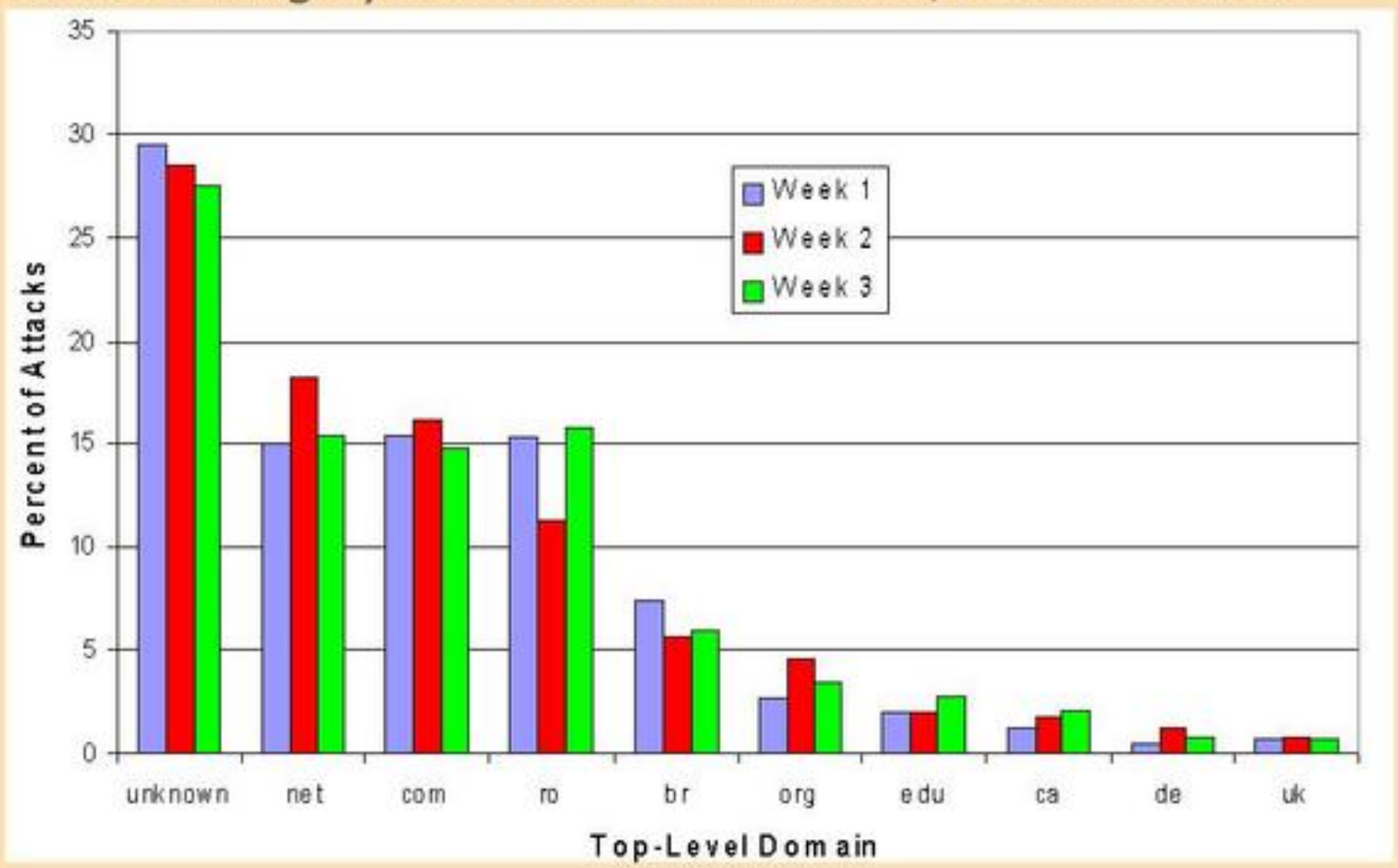
- data:
  - >12,000 attacks against >5,000 targets in 3 weeks
  - ~20-60 attacks occuring at all times
  - 80% of attacks last less than an hour, a few lasted 3 weeks
  - 70% of attacks <1,000 pps, some over 600,000 pps
  - 10-20% of attacks to home machines (cable, dsl, dialup)
  - 5% of attacks target infrastructure (routers, dns servers)
  - (usenix 2001, david,colleen@caida, stefan,geoff@ucsd)

performance myth: DoS attacks

myth:

flooding DoS attacks (randomly spoof src, main attack type)
only affect large commercial sites, are long in duration
and at extremely high rates

- data:
  - romania and brazil have disproportionate number of infected hosts (victims)
  - other domains have roughly same ratio of infected/total machines

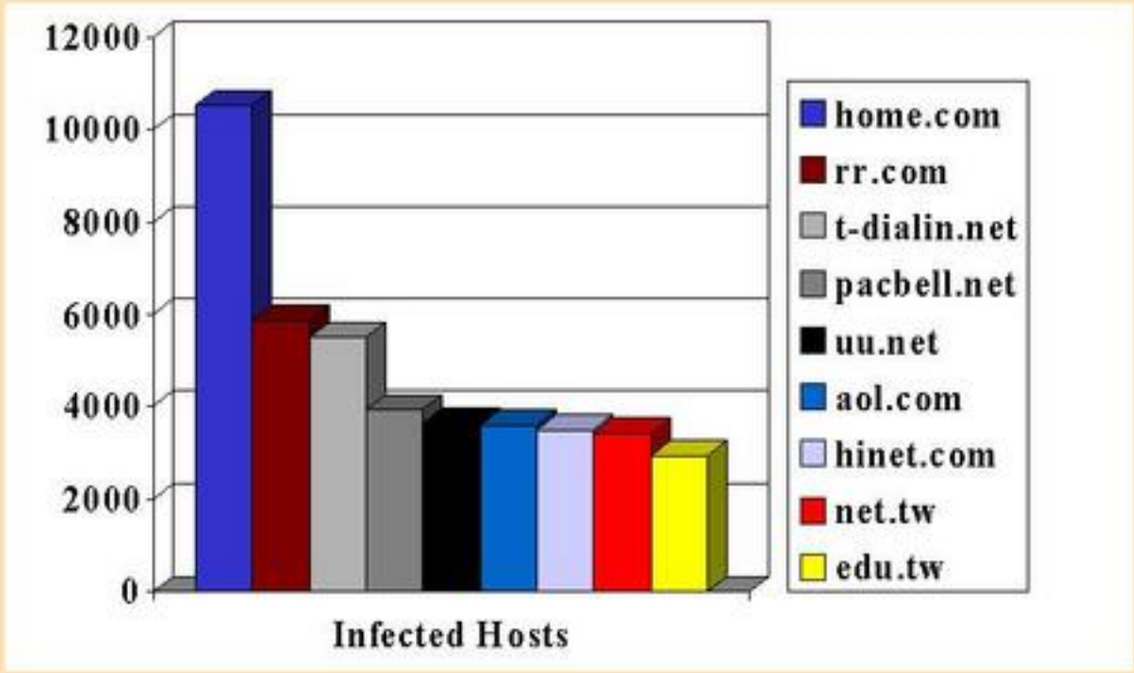


performance myth: worm spread

myth:

worm spread

- data:
  - 40% of all hosts infected (first round CodeRed) lacked reverse DNS records, so we were unable to determine their hostnames
  - ISPs providing connectivity to home & small-business users had the most infected hosts
  - machines maintained by home/small-business users (i.e. less likely to be maintained by a professional sysadmin) are an important aspect of global Internet health

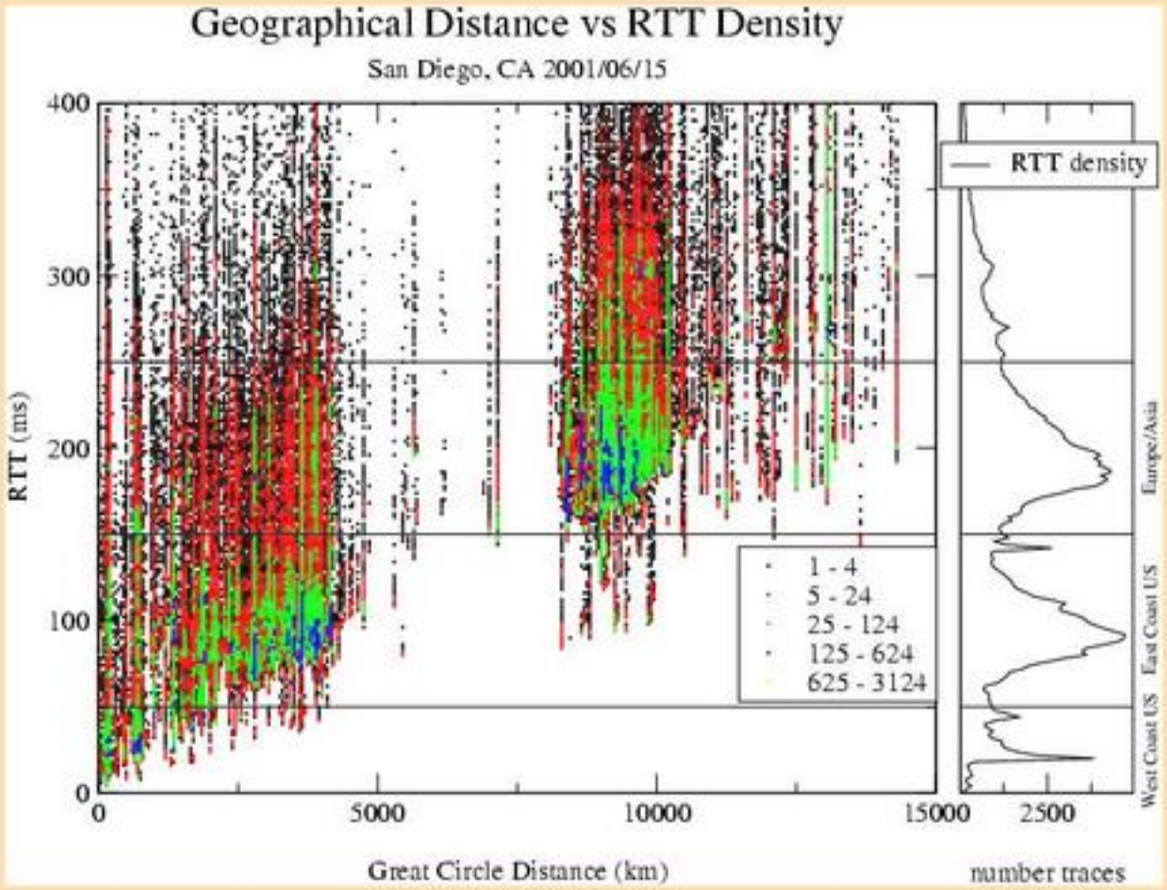


performance myth: geography vs latency

myth:

geography not correlated w/latency

- data:
  - rtt densities from san diego to hosts around world (strong correlation)

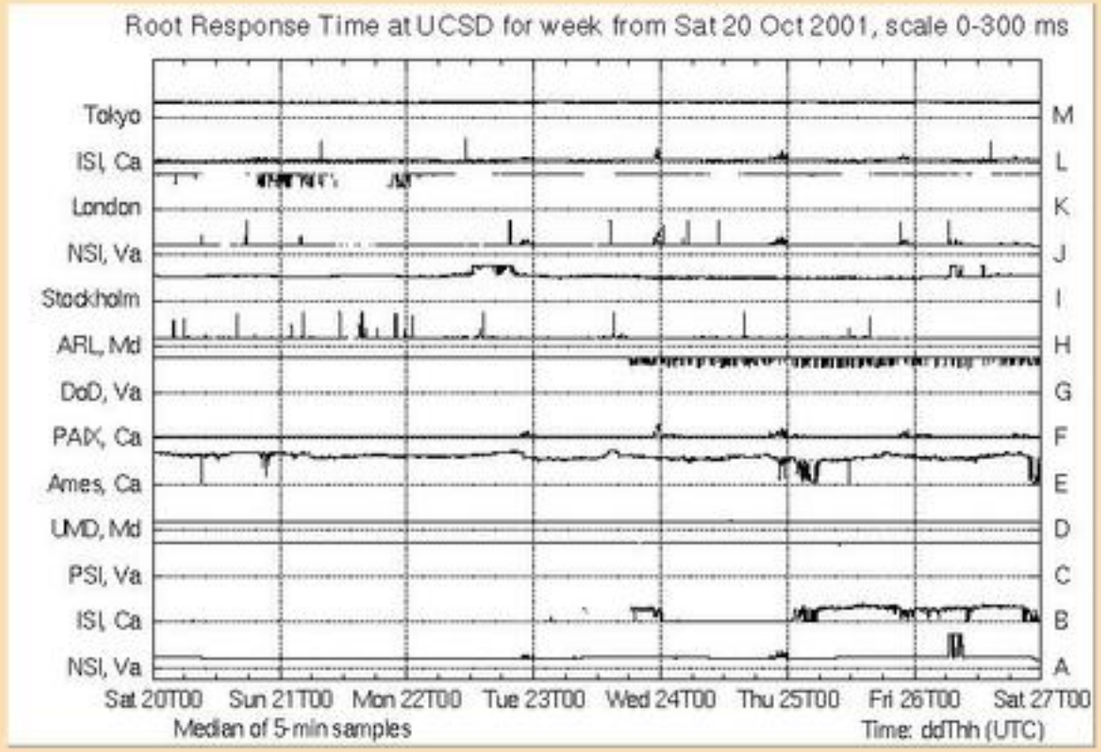


performance myth: root DNS performance

myth:

root DNS system performs well

- data:
  - 8 of the 13 root servers perform well, so users do not notice the poor performance of the other five (gTLD's do somewhat better)



performance myth: DNS performance

myth:

the DNS system performs well

- data:
  - error taxonomy
    - **bogus A queries to root name servers for a few hours at f-root in 2001**
    - **A queries ask for the IP address of a hostname**
    - **not supposed to be 'in theory'**
  - malformed A queries were 14% of the load at F.root
    - **guilty: microsoft: Win2k resolver, viruses (win95/98/nt), macOSX resolver**
    - **asking for the IP address of an IP address**
    - **20% of queries asking for non-existent TLD**
  - lots of internal microsoft names (active directory)
  - lots ending in .local, .localhost, .workgroup, .msft, .domain, etc.
  - hard to track down, nameservers just relay clients queries
  - cannot see back to the actual client that asked the question

performance myth: power of strategic router

myth:

single router can't trash the Internet (`certainly not by accident')

(hint: just need to trash 13 hosts to effectively trash the Internet)

- data (just one example):
  - microsoft's feb 2001 dns woes
  - microsoft's 4 authoritative nameservers visible to world on one subnet (and now all you need is a comma in the wrong place)
  - misconfigured router upstream of that subnet
  - TTL for their names set to 2 hours started timing out of peoples caches
  - query load at the roots started climbing
  - microsoft nameservers don't do negative caching
  - microsoft properties are usually about 6k queries/hour (0%) increased to 25% of the load at f-root

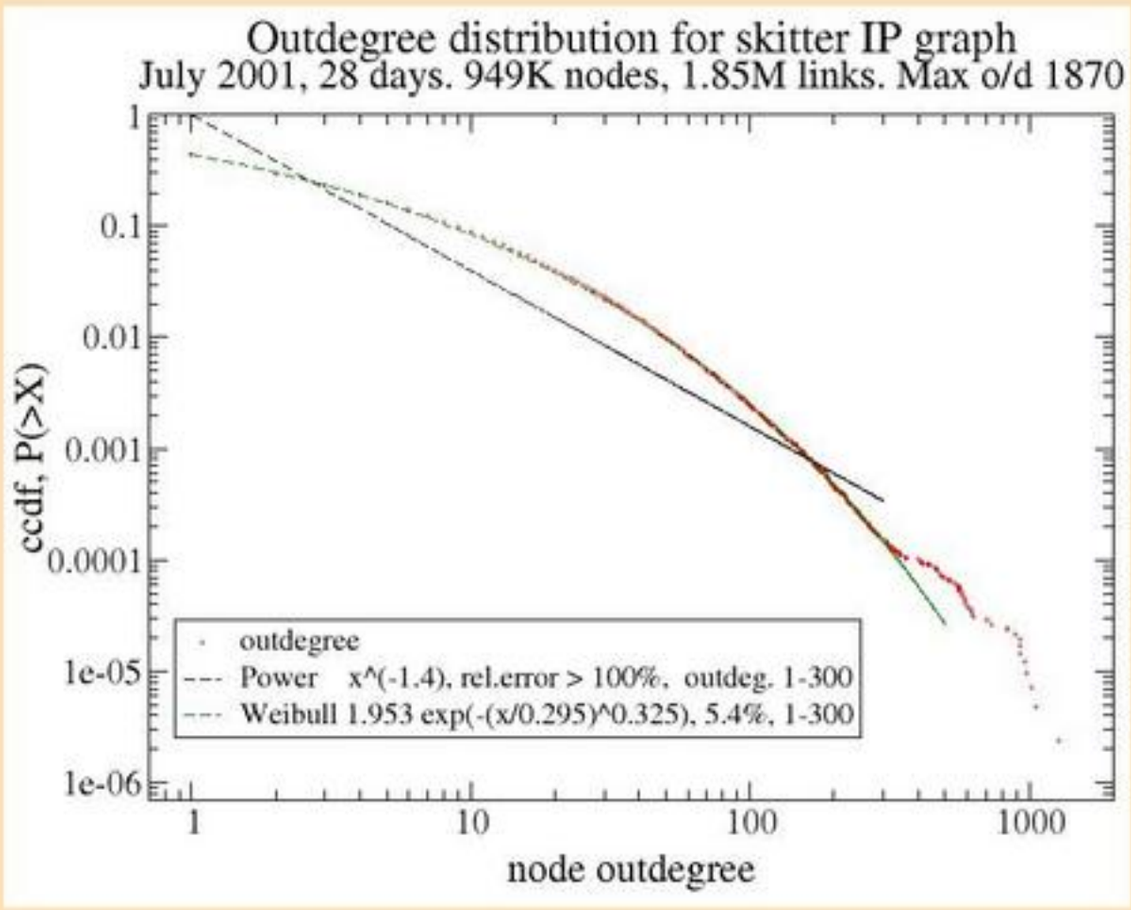
- lesson:
  - prominent site w/DNS problems affects whole Internet
  - cf. 9/11 cnn.com queries to roots were sustainable because of caching
  - this only a tiny piece of the root server workload damage found

topology myth: outdegree distributions

myth:

outdegree distributions follows power law

- data:
  - distribution follows Weibull far better than power law

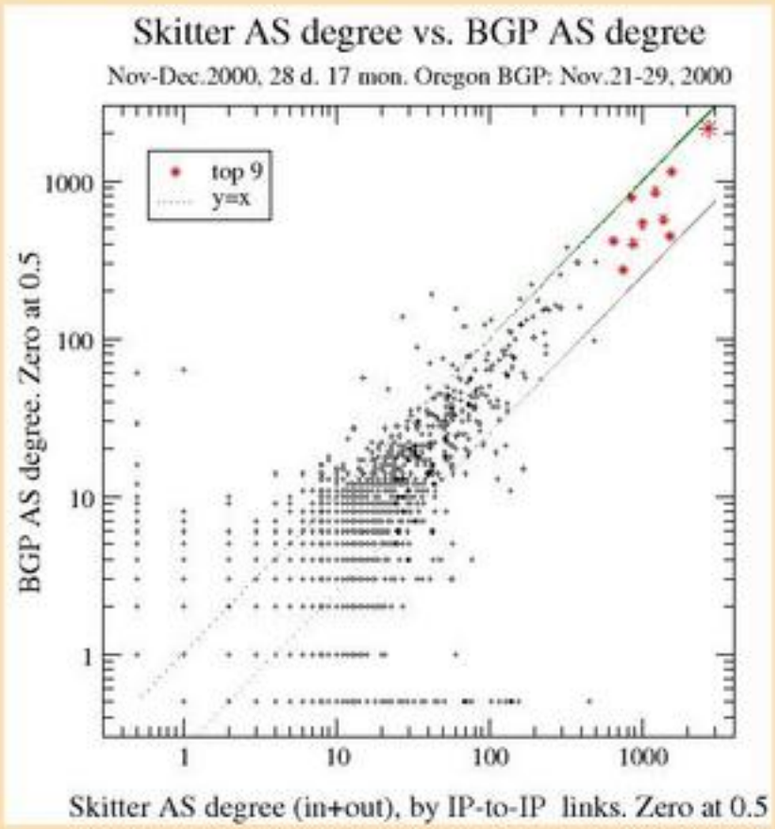


topology myth: routing tables and topology

myth:

routing table data reflects topology

- data:
  - even best avail. inter-domain routing (BGP) data will not reflect true topology
  - (and yet this BGP data is essential to sound macroscopic Internet topology analysis)

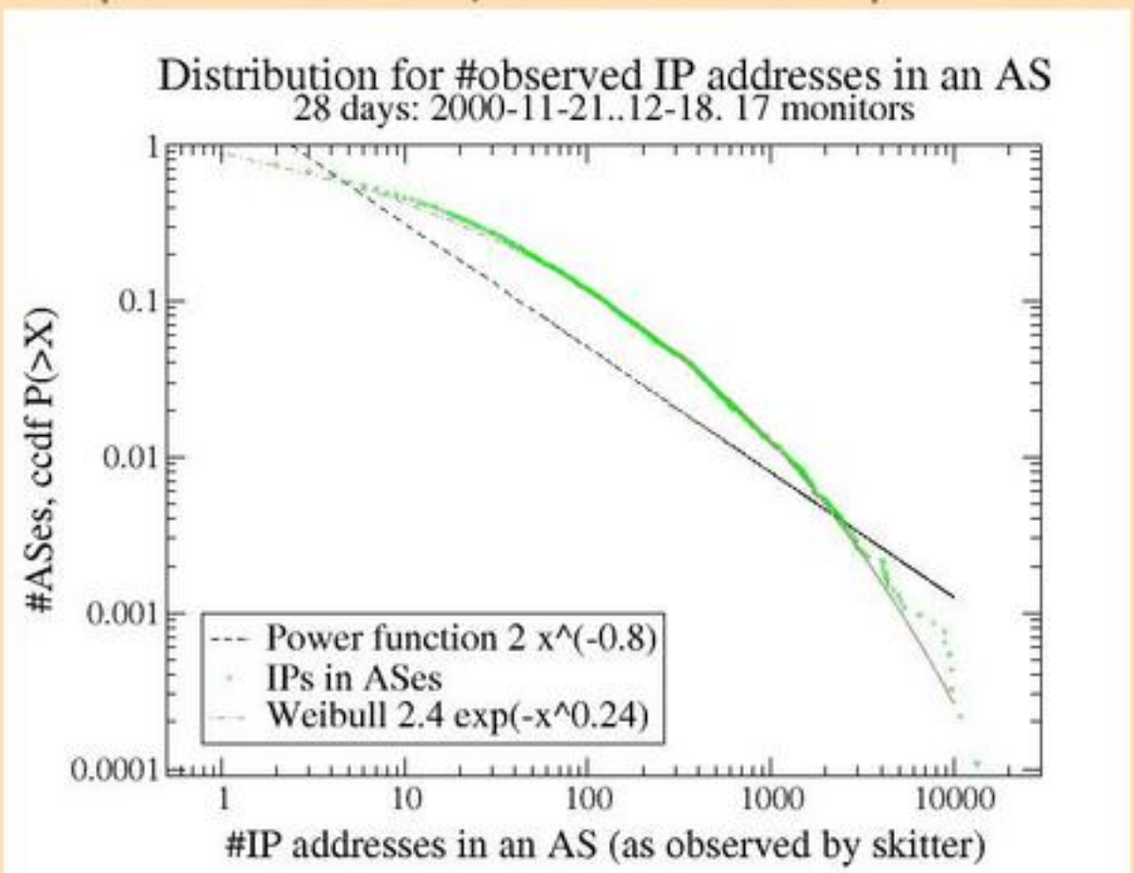


topology myth: Internet object sizes

myth:

Internet object sizes follow power law

- data:
  - Internet graphs are closer to Weibull than to power functions
  - P(X>x) = a exp(-(x/b)^c)
  - decreases faster than power function, slower than exponential

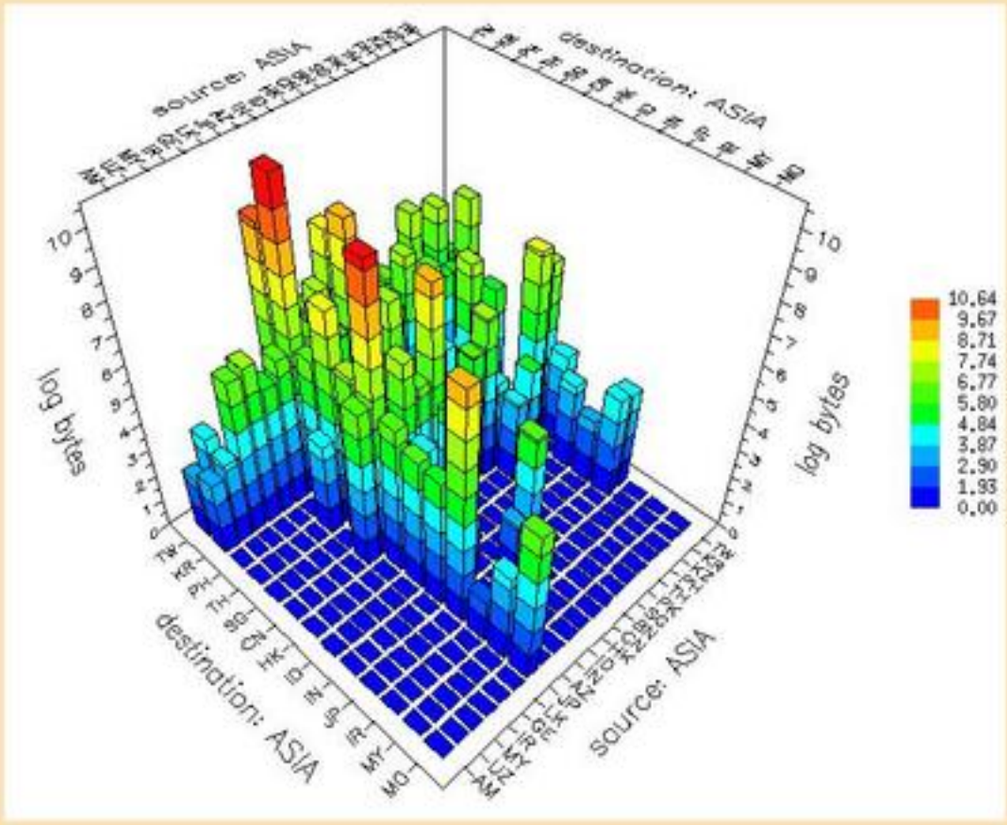


routing myth: intra-country traffic

myth:

intra-country traffic stays there

- data:
  - significant asia-to-asia traffic goes thru san jose
  - includes even same country traffic (e.g., .jp->.jp, .tw->.tw)

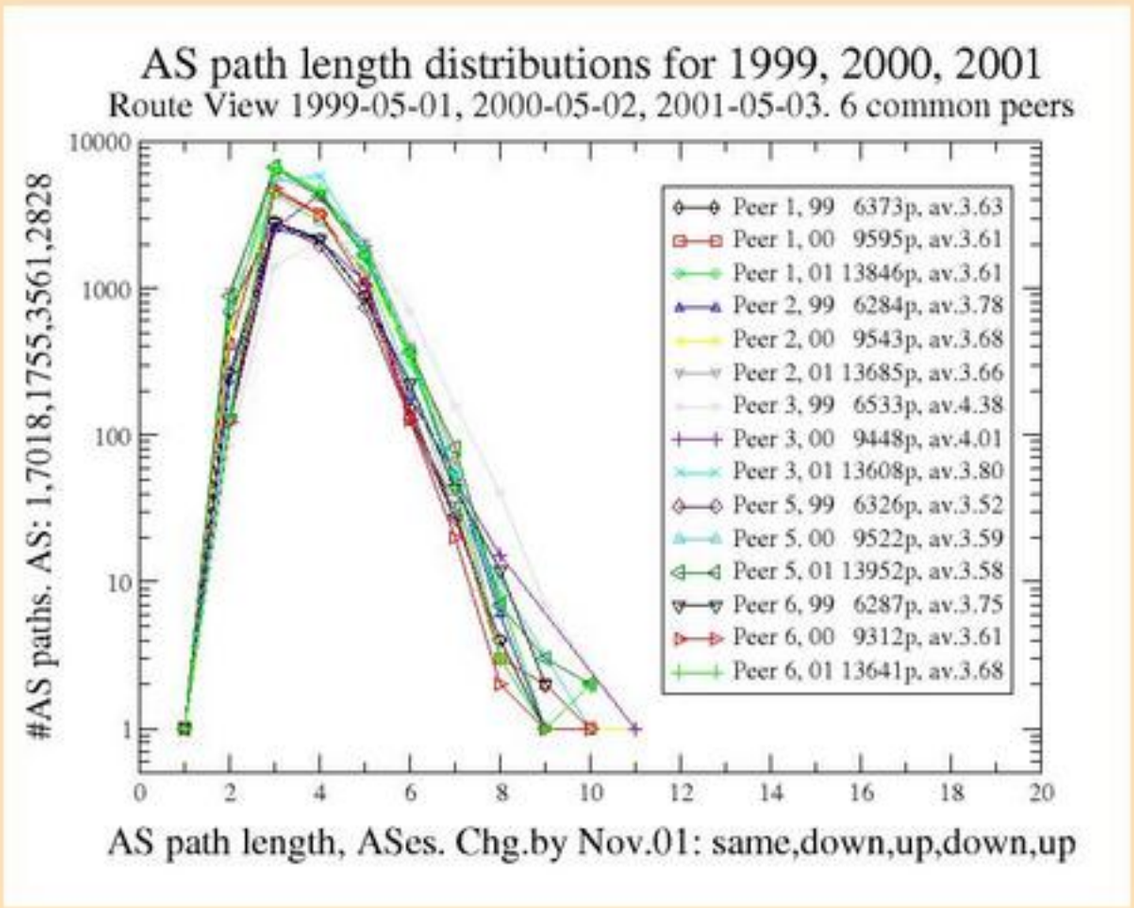


routing myth: AS path length is decreasing

myth:

AS path length is decreasing

- data:
  - since 1999, many AS paths have changed either way
  - average length decreased and increased for many ASes
  - change in the average AS path length is insignificant

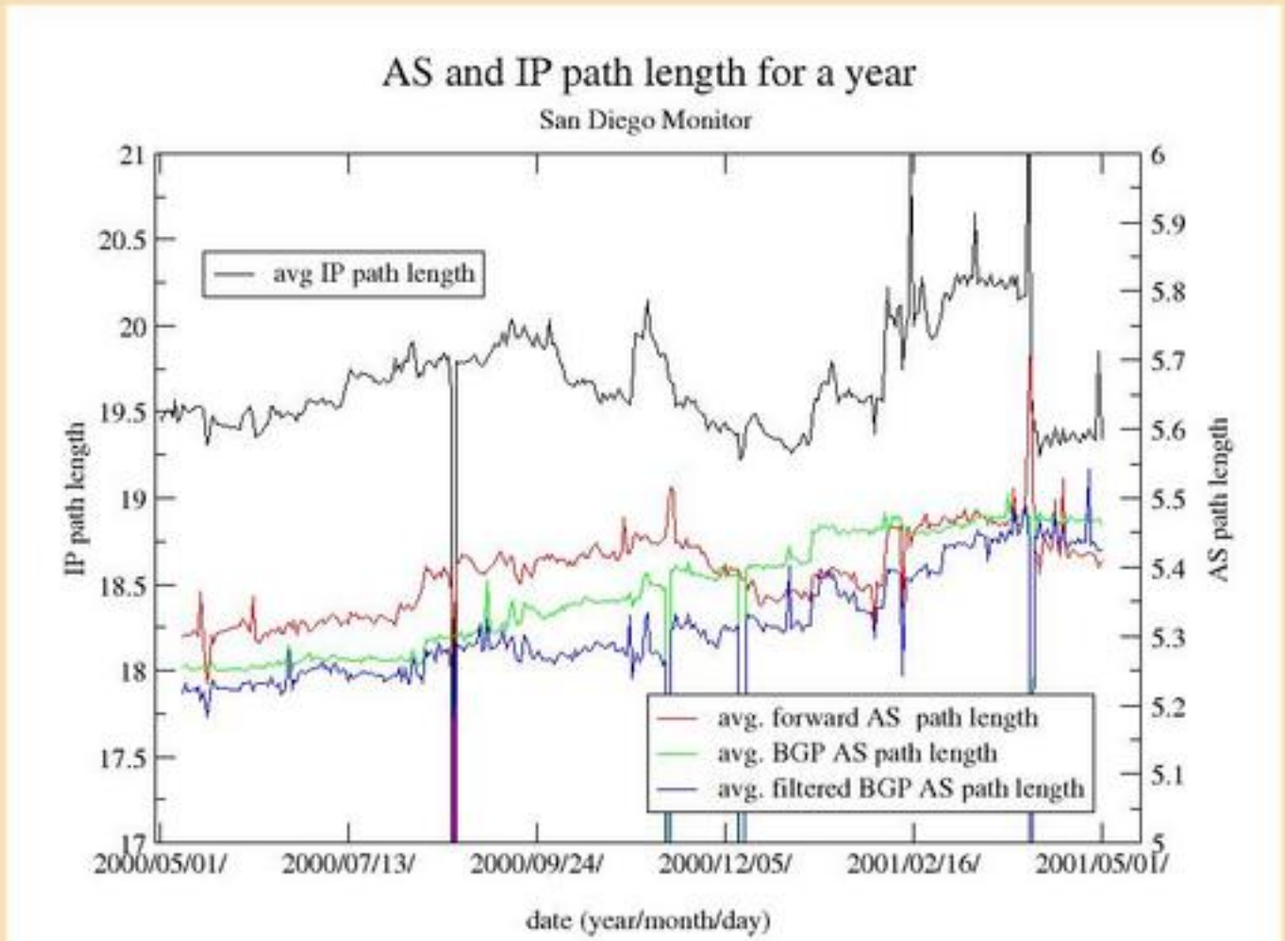


routing myth: AS path length (continued)

myth:

AS path length is decreasing

- data:
  - if anything, it's increasing



routing myths: routing behavior

myth:

causes of growth & instability of routing system
route table growth exponential

- data:
  - global prefixes grew 4% may->nov 01; 37% in nov00-01 (RouteViews)

myth:

peering richness is growing (see previous slide)

- data:
  - link/node ratio (average degree), peering richness, and churn did not significantly change in 2000-2001, although lot of changes within ASes.

routing myths: routing behavior (continued)

myth:

small ISPs & multihoming cause growth and/or churn

- data:
  - since 1998, the share of /24s has stayed between 57% and 58.5%
  - number of non-transit multihomed ASes grew from 35% to 37% in 2000-2001, but their share of global routes remained stable at around 30%.
  - new address announcements & deaggregation of existing prefixes were major sources of new prefixes between nov00-may01
  - most routing instability (w/drawal/reannounce events) in late 2001 contributed by a few .gov networks, developing country telecoms, & major backbone ISPs. although backbone providers routes are relatively stable on per-prefix basis.
  - instability caused in part by deaggregated routes leaking out originating AS, and by relatively short-lived transient announcements. (`small multihomers' contribute negligibly, at least on bi-hourly scale)

Internet myths relevant to engineering

- workload: (besides basic traffic growth fiction)
  - level and nature of fragmented traffic
  - increase in flows as bandwidth grows
  - mice vs elephants
  - private addresses in core
  - prevalence of encrypted passwords
  - applications can be identified (much less controlled)
  - multicast traffic, flows, addressing
- performance:
  - DoS attacks affect only large sites
  - geography not correlated with latency
  - DNS system performs well
  - single router can't trash the Internet
- topology:
  - Internet topologies, object sizes follow power laws
- routing:
  - routing tables reflect Internet topology
  - intra-country traffic stays there
  - AS path length is decreasing
  - small providers and multi-homing (more specifics) cause all the churn

why so many? no real data/measurement...

conclusions

we shed doubt on (too many) commonly assumed Internet myths

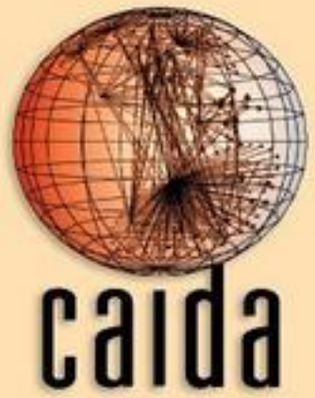
even with use of a number of data sets,
we as a community have quite low integrity
in drawing macroscopic inferences

implication:

the community could make much better use
of our collective intellectual resources if we could
validate ideas against a larger variety of empirical data sets
before investing research and development time and energy on
ideas that attempt to affect the infrastructure

now what?

- `seamless' infrastructure: no such thing (right now)
- measurement tools/architecture
  - well-considered
  - strategically deployed
  - collaboratively maintained
- more operationally relevant research on resulting data
  - feedback into tool design
- correlation among data sources/types, simulation, visualization
- proactive participation
  - top-down (app developers scope constraints)
  - bottom-up (ISP cooperation)
  - middle-matters-more-than-before (dns, mcast, proxies, nats, tunnels, etc)



k r krishnan, `parametric resampling analysis
or traffic measurements for capacity management',
quote from 1999 itc ip seminar:

"kc said `in the absence of data we just press ahead.'
well, we do better than that; we manufacture data.
but we are too refined to call it that,
so instead we call it parametric resampling."

kc
ucsd/sdsc/caida
kc@caida.org
www.caida.org

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

trilogy of action for scientists

for 'seekers of the larger view'

[while video loads...]

- draw together pieces of science and technology to create a system
  - **whether that system is xerography, telegraphy or steam navigation.**
- find the economic feasibility for a new technology
  - **by virtue of a wide grasp of the worlds of man and matter**
- reach harmony through intuition
  - **by meditating on deep knowledge of the field so as to arrive at a new result**
- build a model
  - **a simplified representation of the problem, subject to experimental analysis**
- serve as a science-technologist generalist
  - **who, many times/yr, extracts the missing point out of a complicated situation**
- make decisions or help others make decisions
  - **by imaginative interaction w/alternatives calculated as consequent on those decisions**

-- john archibald wheeler