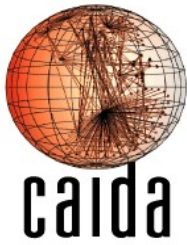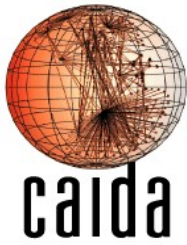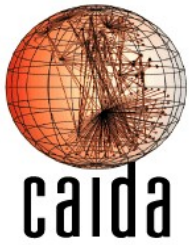# caida *Nuggets*

caida background & history

- since 1997: narrowing the gap between Internet operations and science in face of global privatization

- largely US taxpayer funded (nsf, dhs), plus sponsors

- seek, analyze, communicate salient features of best available data on the Internet

- recent expansion of research agenda into policy and economics

caida recent activities

- data sharing for reproducible research (datcat, PREDICT, DITL, COMMONS)

- infrastructure (hardware and software) upgrades

- dns traffic and vulnerability analyses

- topology measurement, curation, analyses, modeling, simulation

- next generation Internet routing

- security: network telescope, cceid

- policy guidance e.g., ipv4 consumption, blog

Internet Measurement Data Catalog (DatCat)

- First catalog to support thorough indexing and user annotations of Internet measurement data sets.

- DatCat: (http://imdc.datcat.org)

  - facilitates searching for and sharing data among researchers,

  - enhances documentation of datasets via a public annotation system, and

  - advances network science by promoting reproducible research and persistent references.
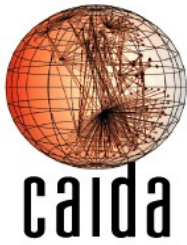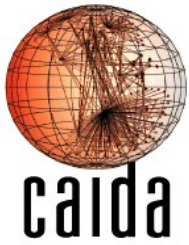
Day in the Life of the Internet (DITL)

- initially, NSF/DNS OARC project to measure as many roots as possible over same interval

- DNS roots (c,f: all; k: 17; m 3; b.osrn, m.orsn)

- collaborators: ISC, NAMES, ORSN, Kaist, Postech, WIDE, Keio, IIT, FIU

- data from: passive hosts, topology probes, telescope

- best test of datcat thus far. 2007 best DITL thus far. sufficient interest -> trying broader scope. need manyeyes! (and supporting s/w)
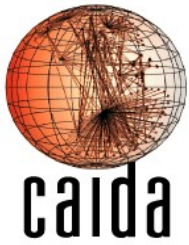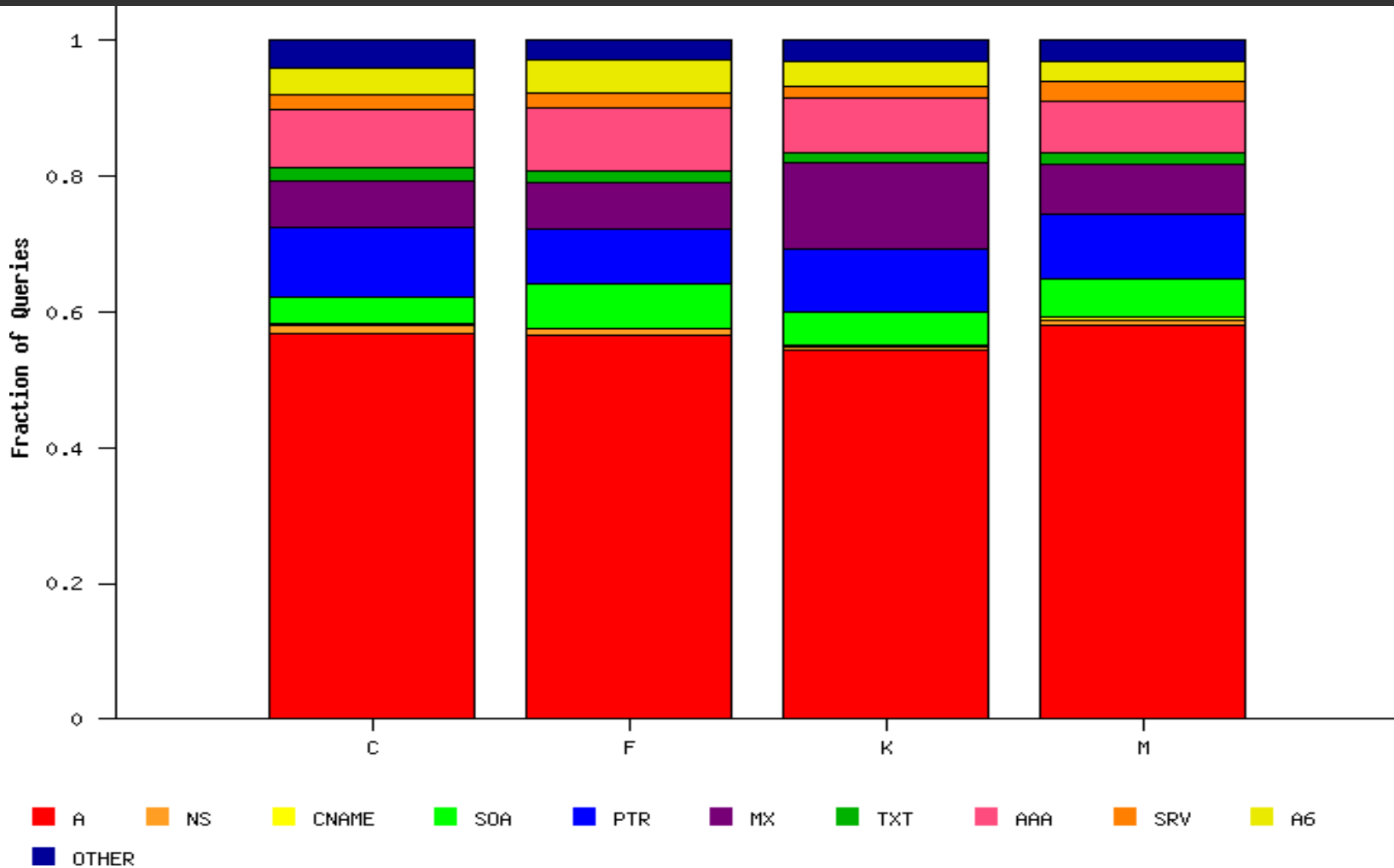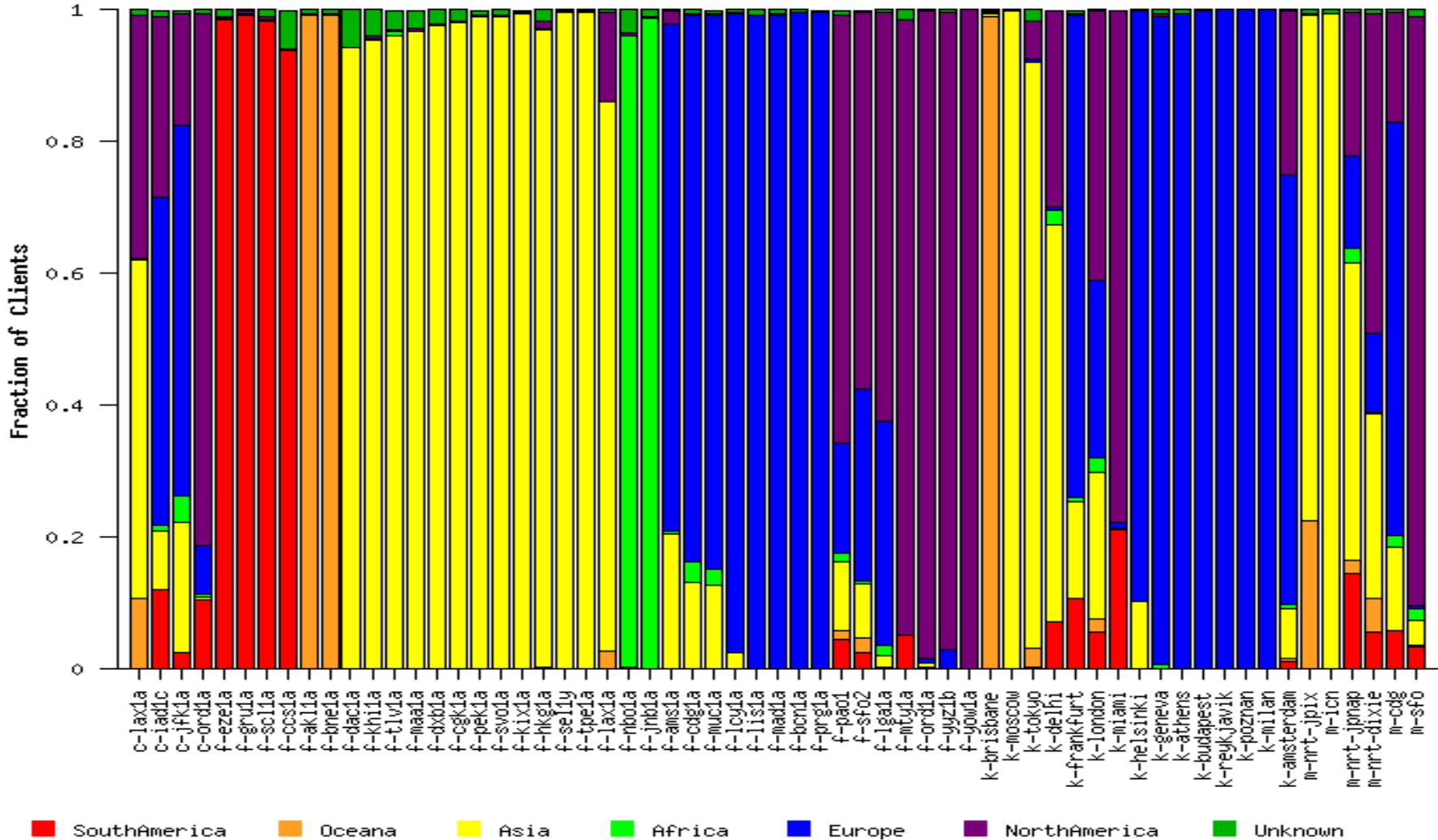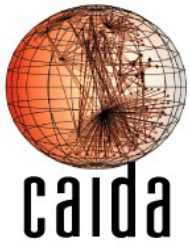
ditl coverage over 48 hours (9jan07)
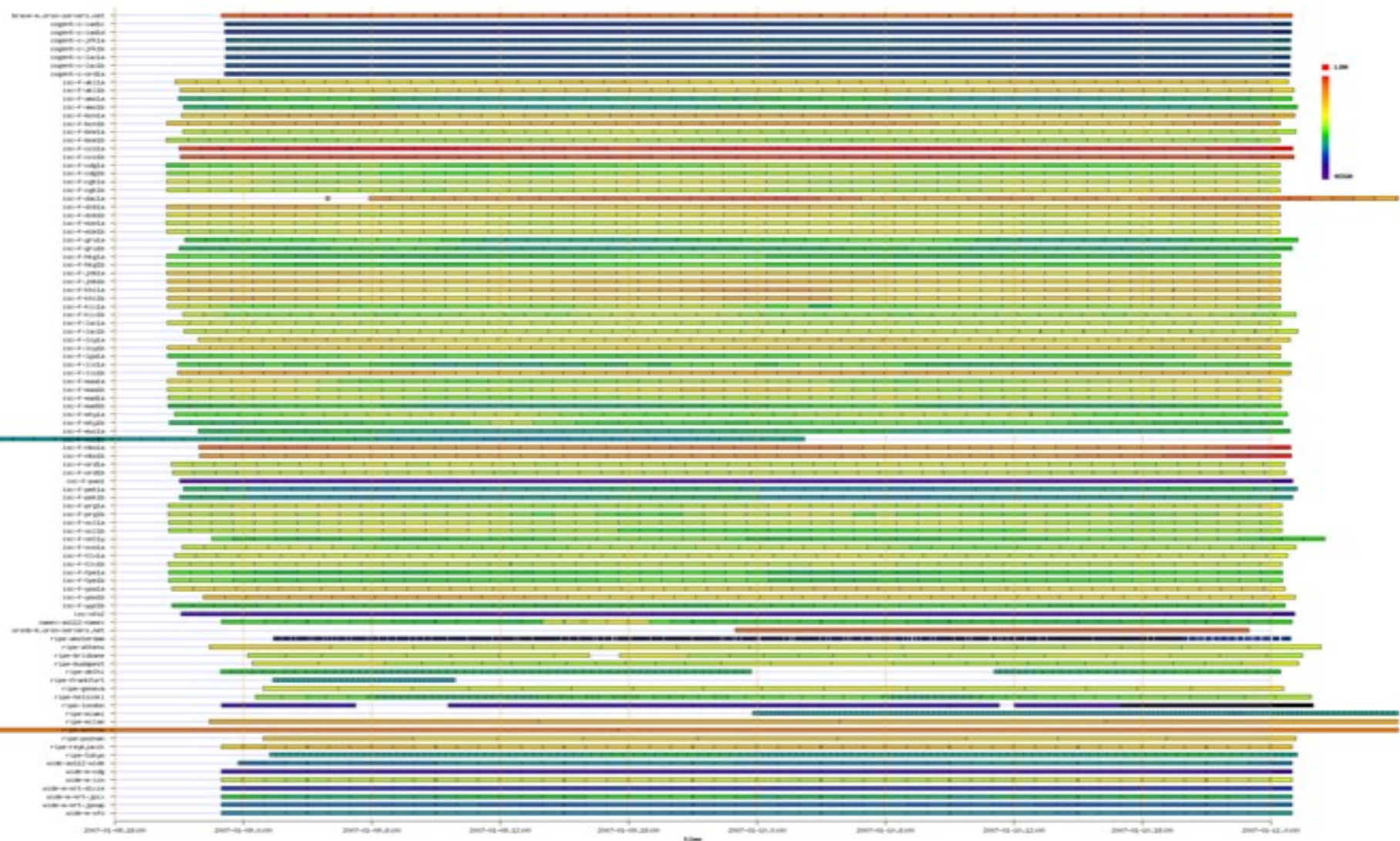
dns queries to 4 root nameservers (9-10 jan 2007, 61 anycast nodes)

dns query sources by region (ditl07)
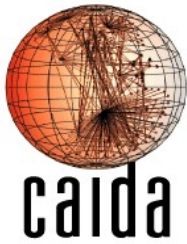
data sharing: muni and community wireless nets

- recognition that commercial data sharing poses persistently daunting if not insurmountable challenges
- which means science is largely stalled
- another approach: COMMONS: cooperative measurement and modeling of open networked systems
- "bandwidth for data trade": exchange overabundance of bandwidth in academic community (I2,NLR) for what the academic community needs most, and has little: technical and economic data on large-scale
- proposition valid on int'l, national, state or local level
- alternative incentive is regulation

measurement software: CoralReef

- CoralReef provides a comprehensive software suite to collect and analyze data from passive Internet traffic monitors, in real time or from trace files.

- The package includes programming APIs for C and Perl, and applications for capture, analysis, and web report generation.

- CAIDA developers maintain this package with the support and collaboration of the Internet measurement community.

- `http://www.caida.org/tools/measurement/coralreef/`

Application bits/s - 1 day
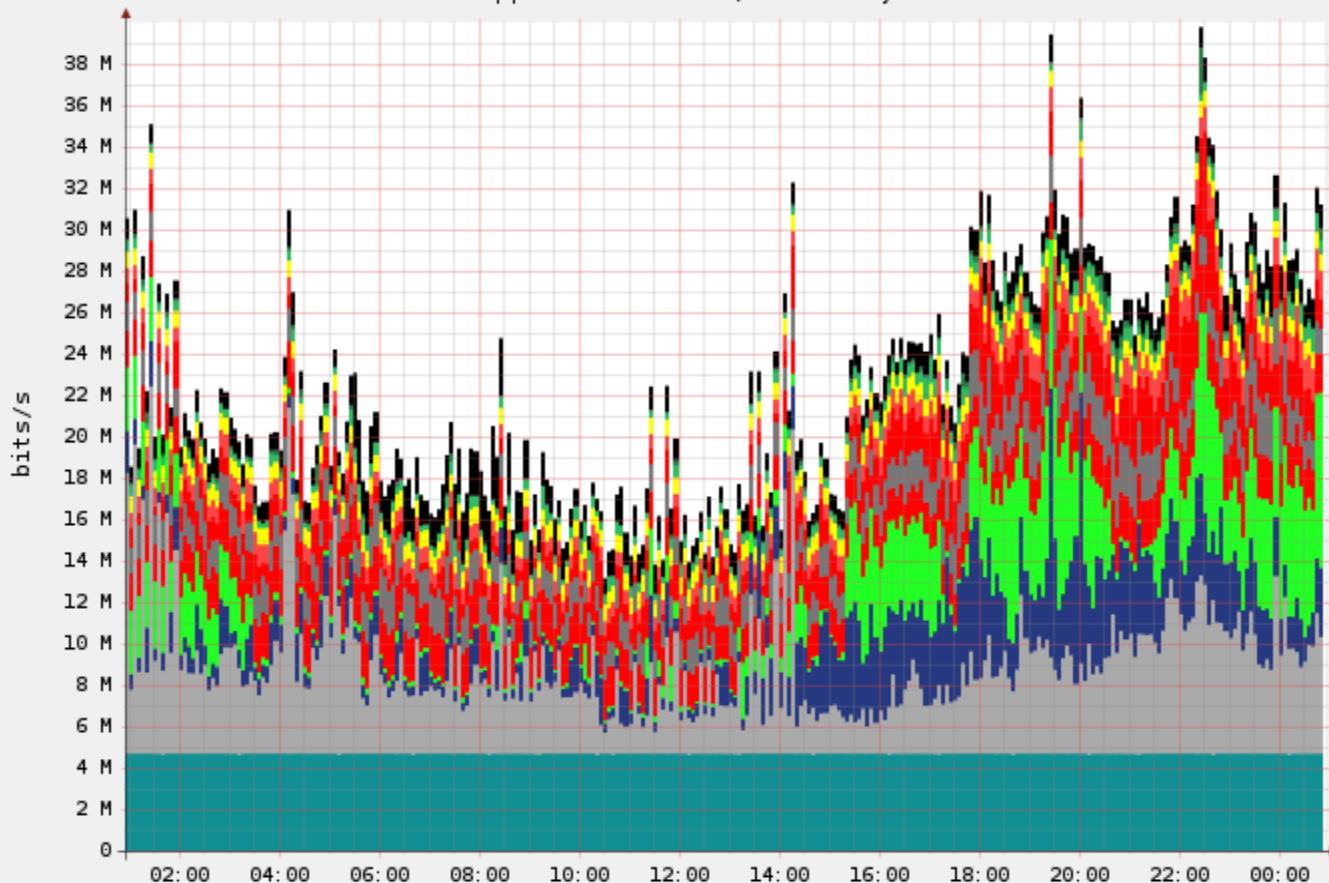
Application	Min	Avg	Max
RCMI_CIAR	4.63M	4.69M	4.70M
UNKNOWN_TCP	1.01M	4.44M	17.74M
IPSEC	344.61k	2.28M	13.08M
SSH	93.50k	2.24M	9.68M
UNIDATA_LDM	1.48M	1.94M	2.39M
UNKNOWN_UDP	916.48k	1.67M	3.05M
HTTP	277.94k	1.58M	5.14M
HTTPS	546.36k	889.34k	1.81M
YAHOO_MESSENGER	792.46k	796.90k	801.28k
SQUID	156.45k	507.31k	1.46M
SMTP	7.35k	135.22k	2.27M
RSYNC	72.05	1.41k	98.64k
other	440.67k	1.20M	3.57M

generated Mon Sep 3 23:45:40 2007 UTC

DNS measurement & analysis

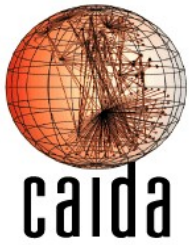- analysis of root server traffic
- dns surveys (open resolvers, cache poisoning, server software)
- anycast measurement, modeling & simulation
- dns measurement software: dsc
- collaboration with NIC chile on measuring .cl
- http://www.caida.org/research/dns/

DNS queries to .cl servers

Queries per minute aggregated by country
seen at .CL anycast cloud
from Tue Jul 3 20:00:01 2007 to Tue Jul 3 21:00:00 2007 (CLT)

Legend — Queries per minute:
0, 1993, 3986, 5979, 7972, 9965, 11958, 13951, 15944, 17937, 19930
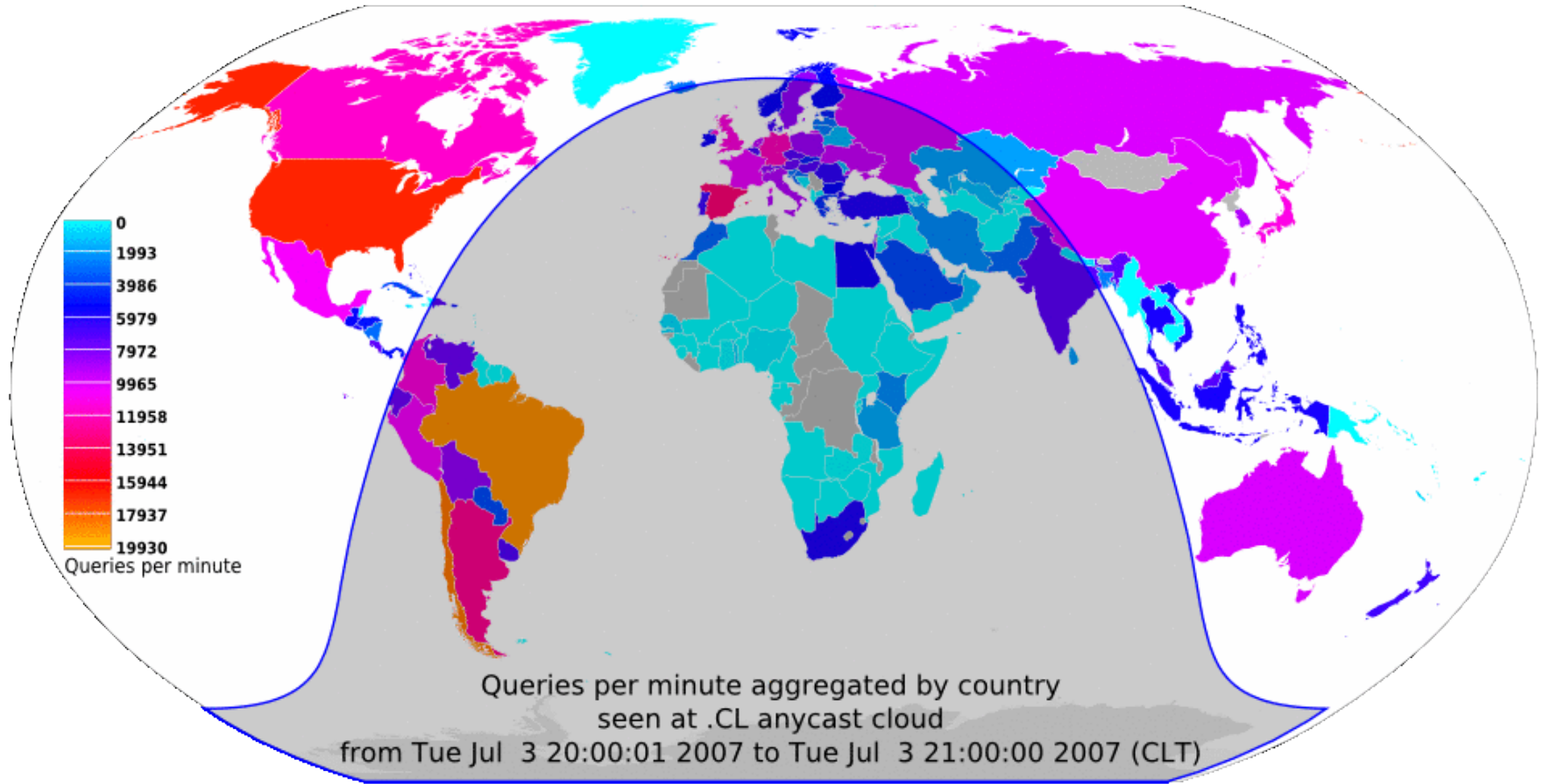
infrastructure: Archipelago (Ark)

- steps toward a community-oriented (active) network measurement infrastructure

- communication and coordination facility to support:

  - time-shared experiments,

  - local and global tuple spaces,

  - storage services,

  - security,

  - extension via Ruby, Perl, C, ?.

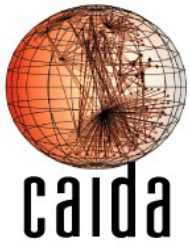 - early apps: skitter/scamper, amplets, dns open resolver probes.

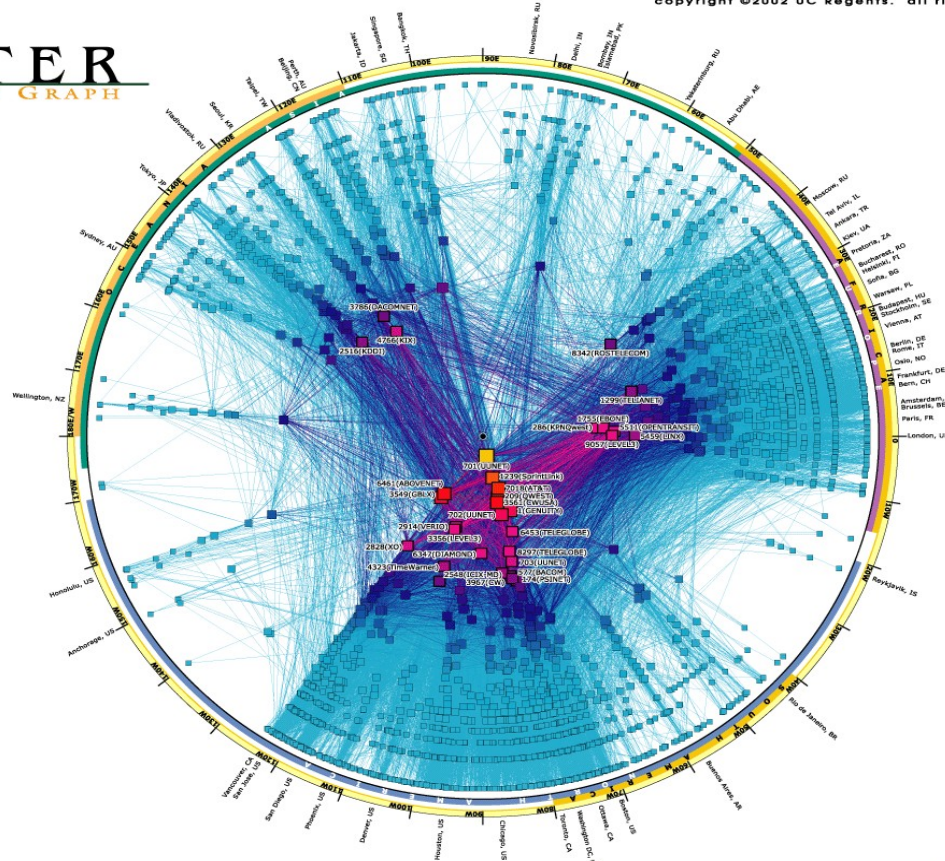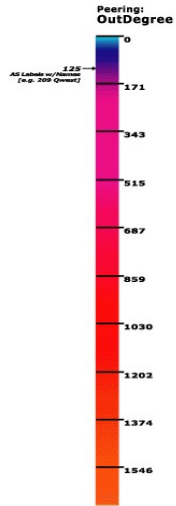# Internet Topology Data Kit (ITDK)

- contains:
  - post-processed skitter+BGP Internet topology data
  - raw files (with IP addresses!), scripts, statistics
  - one of the most complete and exhaustive Internet IP/router-level topology data sources available

  - `http://www.caida.org/tools/measurement/skitter/idkdata.xml`

http://www.caida.org/

skitter-derived AS topology graphs

- daily summaries of the Internet AS-level topology as seen by skitter

- format listing of all AS links comprising the AS graph

- search terms: **skitter topology as adjacencies**

- `http://www.caida.org/tools/measurement/skitter/as_adjacencies.xml`

skitter-derived router-level topology

- one-time snapshot of the Internet router-level topology as seen by skitter and iffinder.

- listing of all router links comprising the router graph.

- free download.

- search terms: **skitter router-level topology**
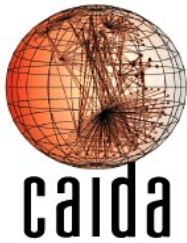
AS topology comparisons

- mapped all commonly considered topology metrics to unifying set of well-known statistical characteristics of network topologies.
- compared major sources of AS-level topology data in this framework
  - skitter (traceroute)
  - BGP tables
  - BGP updates
  - WHOIS
- found that joint degree distribution (JDD) plays a definitive role:
  - peculiarities of data sources explain differences in JDDs across topologies,
  - differences in JDDs, in turn, explain differences in all other characteristics.

AS topology comparisons (continued)

- most exhaustive and unifying source of AS topology:
  - extracted topology graphs,
  - calculated all important topology characteristics (both plots and associated datasets!), and
  - scripts used to calculate them.
- free download
- URLs
  - data: http://www.caida.org/analysis/topology/as_topo_comparisons/
  - paper (short): http://www.caida.org/publications/papers/2006/as_topology/
  - paper (long): http://www.caida.org/publications/papers/2005/tr-2005-02/

AS relationships

- established a new standard for AS relationship inference
- ran experiment that accurately inferred approximately 90% of all peering links.
- validated the results extensively with ~40 ASes.
- weekly snapshots (starting 2004) of AS-level topology augmented with AS relationships
- URLs:
    - data: http://www.caida.org/data/active/as-relationships/index.xml
    - paper: http://www.caida.org/publications/papers/2006/as_relationships_inference/
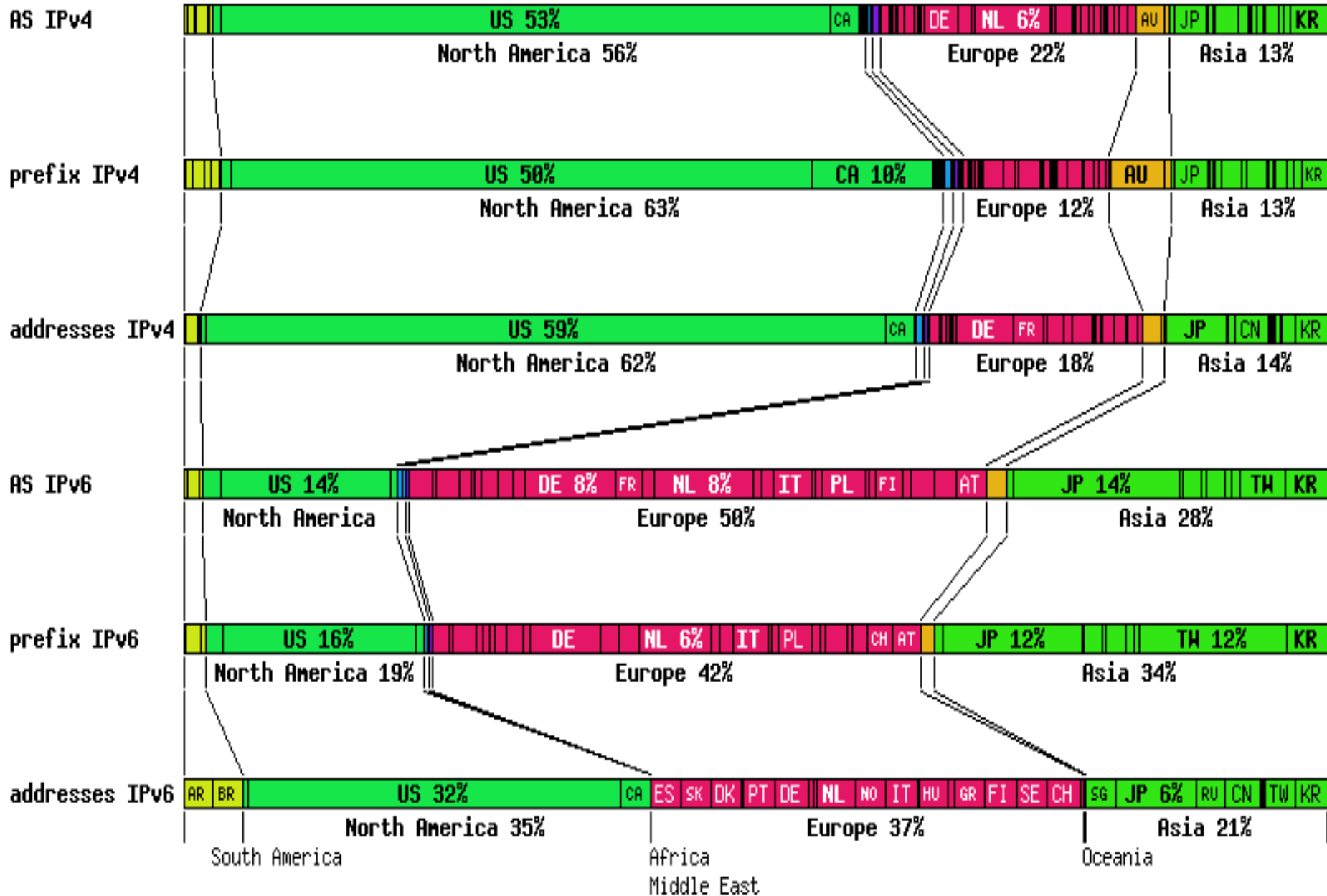
AS Ranking

- rank observed ASes based on hierarchy induced by inferred AS relationships.
- the higher the rank of an AS, the greater the IP address space advertised by the AS and its customer cone (generally)
- http://as-rank.caida.org/

(2005)

percentage of metric controlled by a country/continent

AS IPv4
US 53% | CA | DE | NL 6% | AU | JP | KR
North America 56% | Europe 22% | Asia 13%

prefix IPv4
US 50% | CA 10% | AU | JP | KR
North America 63% | Europe 12% | Asia 13%

addresses IPv4
US 59% | CA | DE | FR | JP | CN | KR
North America 62% | Europe 18% | Asia 14%

AS IPv6
US 14% | DE 8% | FR | NL 8% | IT | PL | FI | AT | JP 14% | TW | KR
North America | Europe 50% | Asia 28%

prefix IPv6
US 16% | DE | NL 6% | IT | PL | CH | AT | JP 12% | TW 12% | KR
North America 19% | Europe 42% | Asia 34%

addresses IPv6
AR | BR | US 32% | CA | ES | SK | DK | PT | DE | NL | NO | IT | HU | GR | FI | SE | CH | SG | JP 6% | RU | CN | TW | KR
North America 35% | Europe 37% | Asia 21%
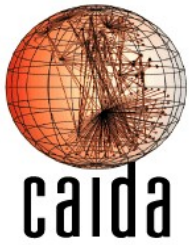
South America

Africa
Middle East

Oceania

AS Taxonomy

- first attempt to classify all ASes.
- inferred AS relationships, WHOIS records, other AS attributes into a simple machine learning algorithm to classify ASes into: large and small ISPs, customers, universities, IXPs, and NICs.
- one-time snapshot of annotated AS-level Internet topology
- best dataset/paper award at PAM
- free download
- URLs:
  - data: http://www.caida.org/data/active/as_taxonomy/
  - paper: http://www.caida.org/publications/papers/2006/revealingas/

Realistic Topology Generation: dK-series

- deep insight into interdependencies among topology metrics
- proposed set of metrics that define all others, even yet unknown ones.
- end few-decade quest among network and graph theorists
- provide framework to formalize connections between:
  - statistical ensembles of networks exhibiting a given set of target properties and models explaining how such networks might have evolved,
  - local and global properties of network topologies, and
  - continuous and discrete representations of complex systems.
- network topology generator strategy capable of constructing synthetic topologies that reproduce properties of target (observed) topologies with arbitrary degree of accuracy.
- applies to broad spectrum of scientific research including, e.g., computer science (both networking and theory), physics, biology, social, economic, and political sciences.
- http://www.caida.org/publications/papers/2006/dk_series/

dK-annotations (in progress)

- extends *dK*-series with semantics of nodes/links
  - AS-level Internet: nodes of different AS types, links of different AS relationships
  - router-level Internet: links of different bandwidth, latency, packet loss, etc.
- first topology generator capable of reproducing *annotated* network topologies.

Internet Evolution (in progress)

- first Internet growth model to combine following three properties:
  - realistic – based on a formalization of Internet economics,
  - analytically tractable, and
  - all its external parameters are measurable.
- predicts and explains not only the observed power laws, but also the exact value of the exponent of the power-law node degree distribution $\gamma \sim 2.1$, thus "closing a loop."
- prediction of value of exponent of the power-law node degree distribution sheds new light on an area that intersects with many different complex networks.
- other contributions of the model include:
  - an attempt to derive standard, i.e., linear, preferential attachment from economic realities of the AS-level Internet, and
  - predicts that topological awareness of Internet players leads to super-linear preference and consequently to deviations from power laws to condensed monopoly states.
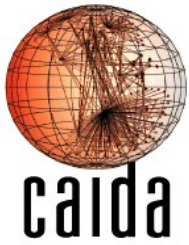- http://arxiv.org/abs/cs.NI/0608058

Name-independent (flat) Routing (in progress)

- growing understanding that hierarchical routing cannot scale since it encodes topological location information in node addresses, which thus need to change with topology change

- algorithms that encode topology information in node addresses *name-dependent*; those that allow any (flat) node addresses are *name-independent.*

- through implementation, we found that state-of-the-art name-independent compact routing algorithms ones incur only minor routing table size increase, compared to name-dependent.

- however, the average stretch (ratio of path lengths produced by an algorithm to the shortest path lengths) increase is significant (from almost no stretch (almost 1), to almost 2). for some applications, such increased stretch is not a problem, for others it is.

- consequently, we plan to:
  - find name-independent algorithms with small stretch and estimate how their stretch scales with network size; or
  - find alternatives to traditional compact routing.

Routing using hidden metric spaces (in progress)
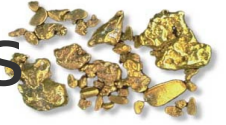
- recognition that the ultimate problem with routing scalability is the *updates*
- begs question: can we route without updates? (it's been done before)
- further cause for optimism: structure of observed complex networks (strong clustering, specific power laws) maximizes their navigability in a hidden metric space
- still need to figure out why!

Internet AS map (BGP)
$\gamma = 2.2, \alpha = 5.0$
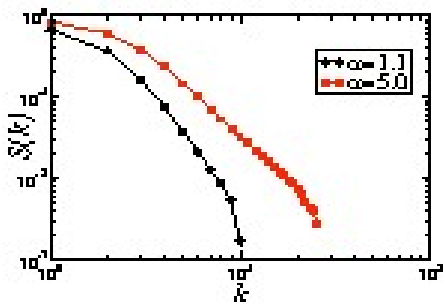$\gamma = 2.5, \alpha = 5.0$
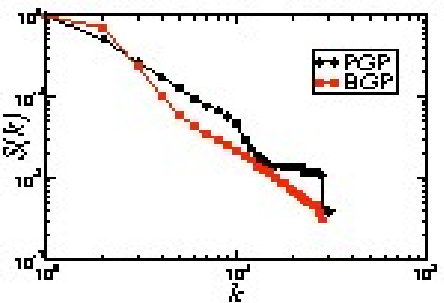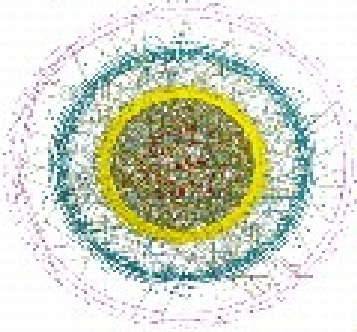$\gamma = 3.0, \alpha = 5.0$
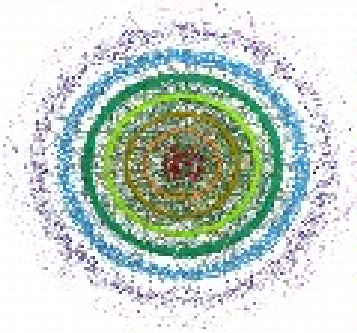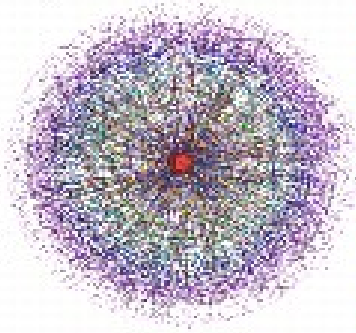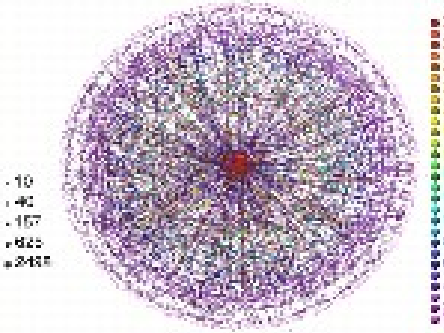
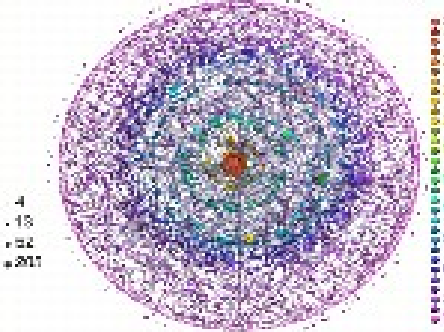Social network (PGP)
$\gamma = 2.2, \alpha = 1.1$
$\gamma = 2.5, \alpha = 1.1$
$\gamma = 3.0, \alpha = 1.1$
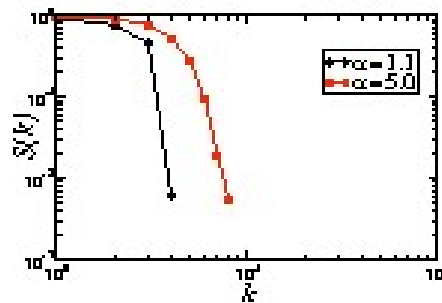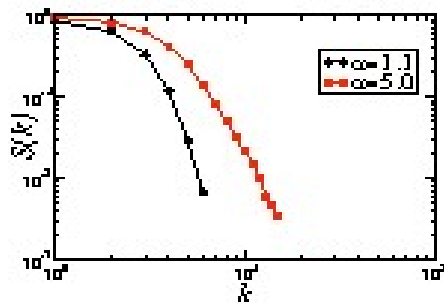
Security: UCSD Network Telescope

- The Backscatter-2006 Dataset
  - This dataset contains information useful for studying denial-of-service attacks.
- Code-Red Worms Dataset
  - This dataset contains information useful for studying the spread of the Code-Red version 2, and CodeRedII worms.
- Witty Worm Dataset
  - This dataset contains information useful for studying the spread of the Witty worm.
- `http://www.caida.org/data/passive/network_telescope.xml`
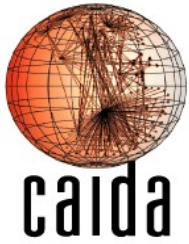
Internet policy guidance: address exhaustion

- IPv4 address depletion
- IPv6 uptake
- what to do about it.. (see EOT talk)
- 2005 ARIN meeting: 'apocalypse then' talk
- 2007: visualizing Ipv4 and Ipv6 reachability with donations of data from isi.edu
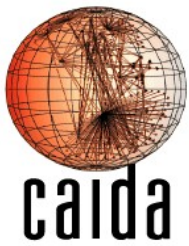
Internet policy guidance: address exhaustion

(play movie version!)

[ping data from isi.edu; poster by Duane Wessels@TMF]