caida priorities: 2006-2008

wide/cnrs workshop
09 feb 2006

kc@caida.org

outline

- list of projects (we cannot complete them all)
- goal: explain where we've been, how we've done, where we're going
- goal: gather feedback from you

too much to do..

- global consumption of IP addresses
- macroscopic topology measurement
- dns analysis
- security analysis
- “day-in-life-of-Internet”
- economics of provisioning
- scalable routing for future networks

1. global consumption of IP addresses

Goal. Provide empirical data and analysis to support informed political decisions governing IPv4 and IPv6 addresses allocation, advertisement, and utilization rates.

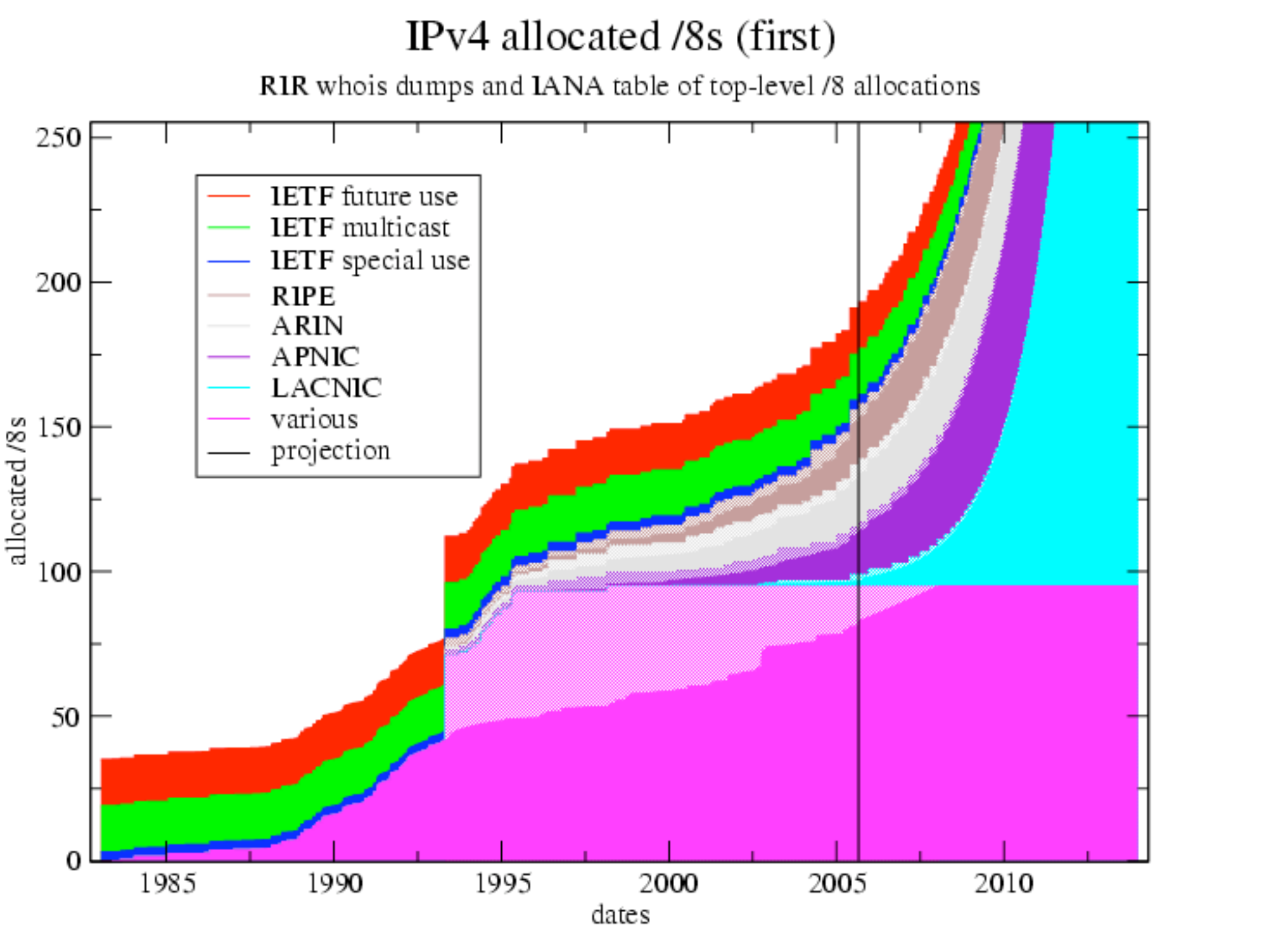
1. gave talk at ARIN Oct 2005 meeting: "apocalypse then".

http://www.caida.org/outreach/presentations/2005/arin/arin200510.pdf

2. working on web site, additional stats, wiki

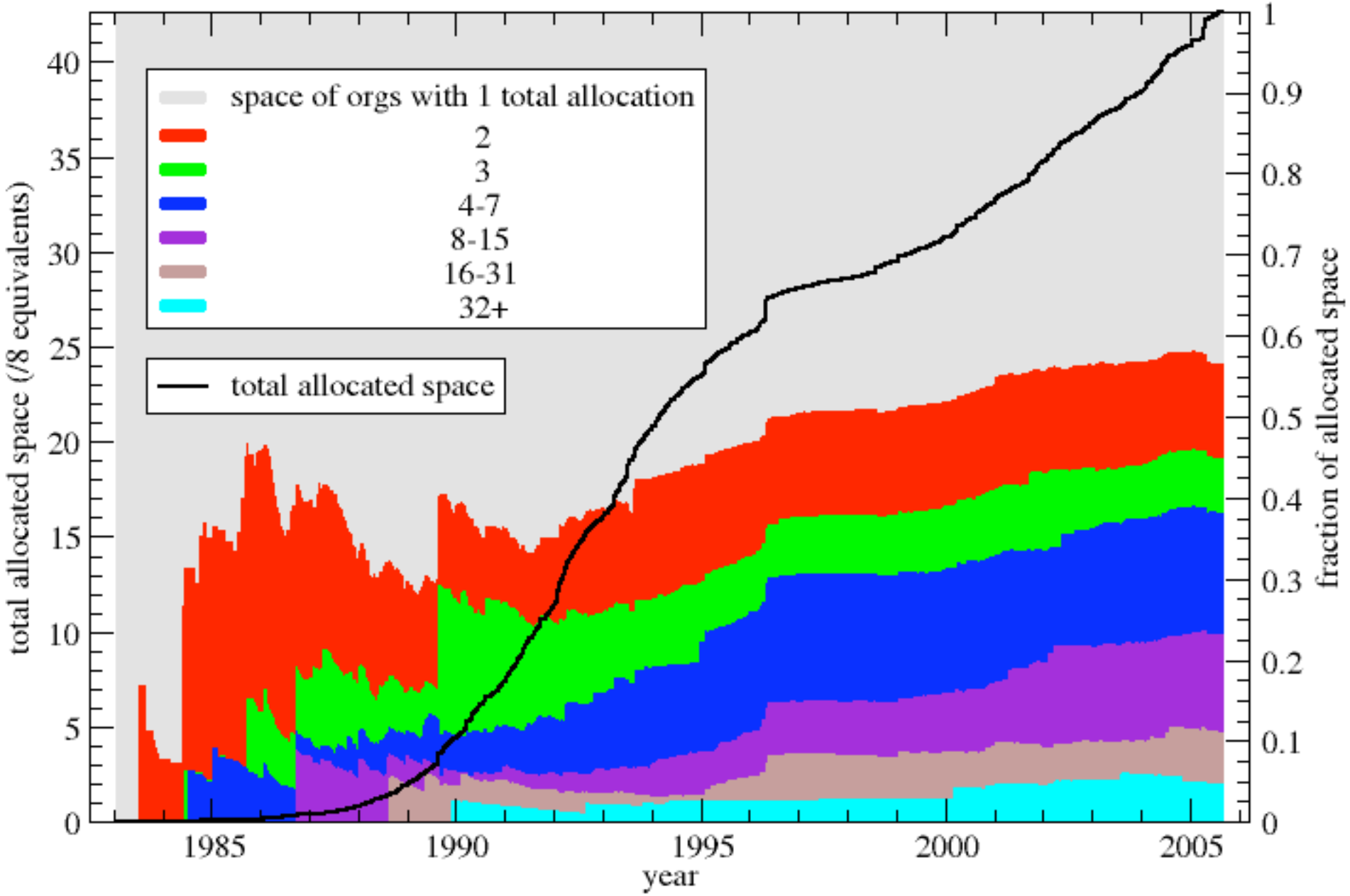
http://www.caida.org/analysis/id-consumption/ipv4/concentration.xml

3. support workshop on future scenarios?


IPv4 allocated /8s (first)
RIR whois dumps and IANA table of top-level /8 allocations
IETF future use
IETF multicast
IETF special use
RIPE
ARIN
APNIC
LACNIC
various
projection
allocated /8s
250
200
150
100
50
0
1985
1990
1995
2000
2005
2010
dates

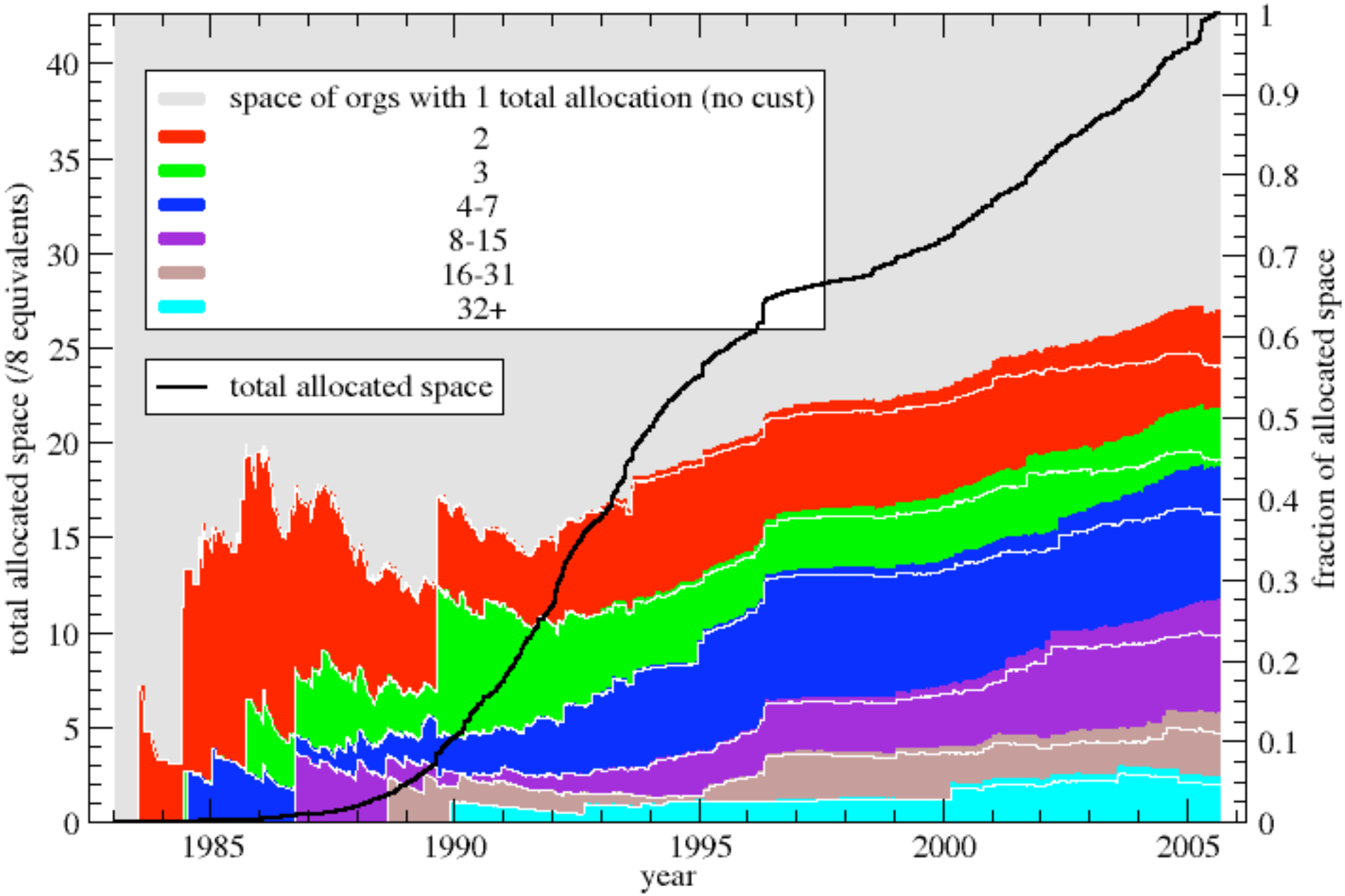
Breakdown by Num Allocations per Organization of ARIN IPv4 Space

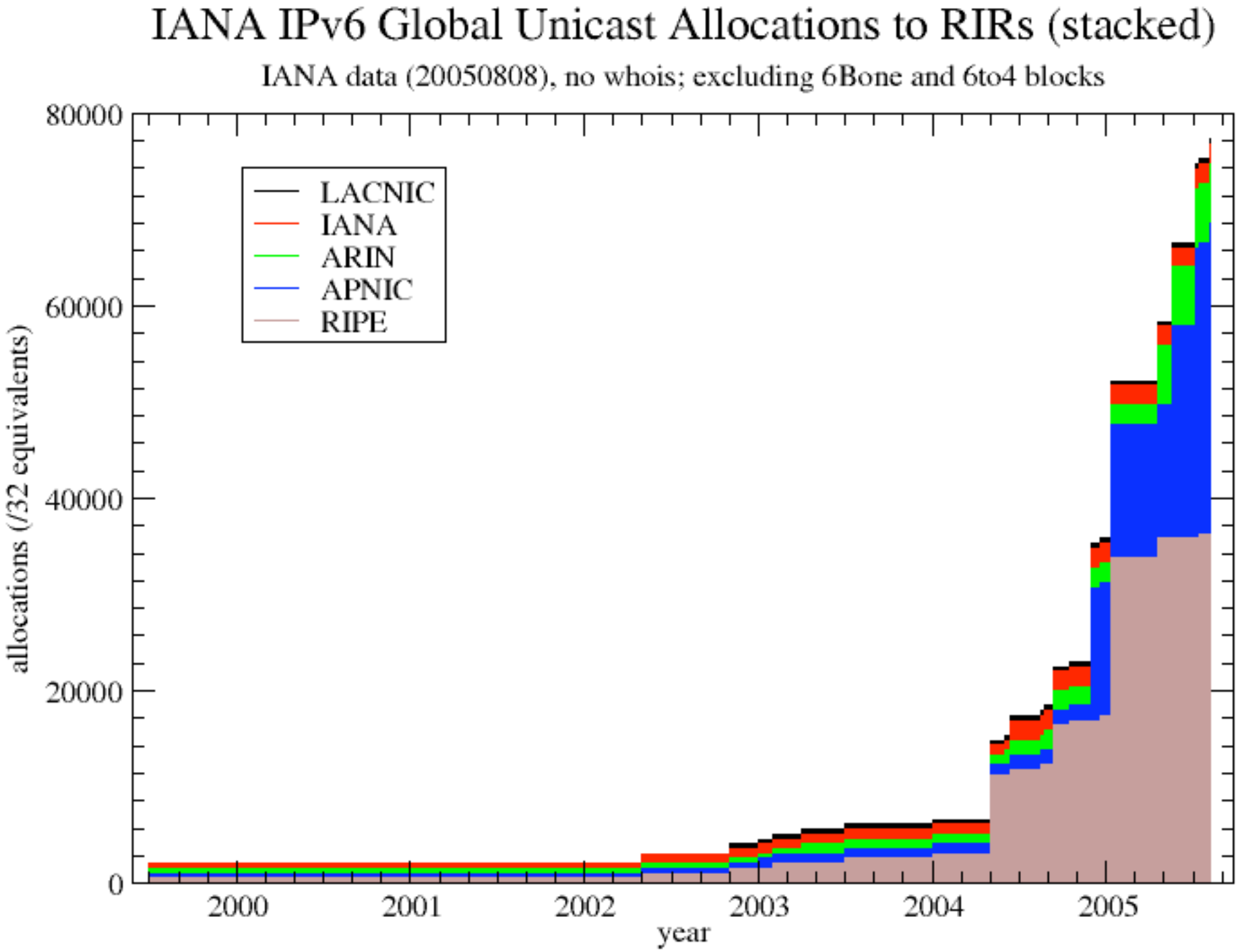
ARIN whois data (20050831); excluding DoDNIC, JPNIC, and pre-RIR /8 allocations; stacked plot; v4



Breakdown by Num Allocations per Organization of ARIN IPv4 Space

ARIN whois data (20050831); excluding DoDNIC, JPNIC, and pre-RIR /8 allocations; stacked plot; v4




IANA IPv6 Global Unicast Allocations to RIRs (stacked)
IANA data (20050808), no whois; excluding 6Bone and 6to4 blocks
LACNIC
IANA
ARIN
APNIC
RIPE
allocations (/32 equivalents)
80000
60000
40000
20000
0
2000
2001
2002
2003
2004
2005
year

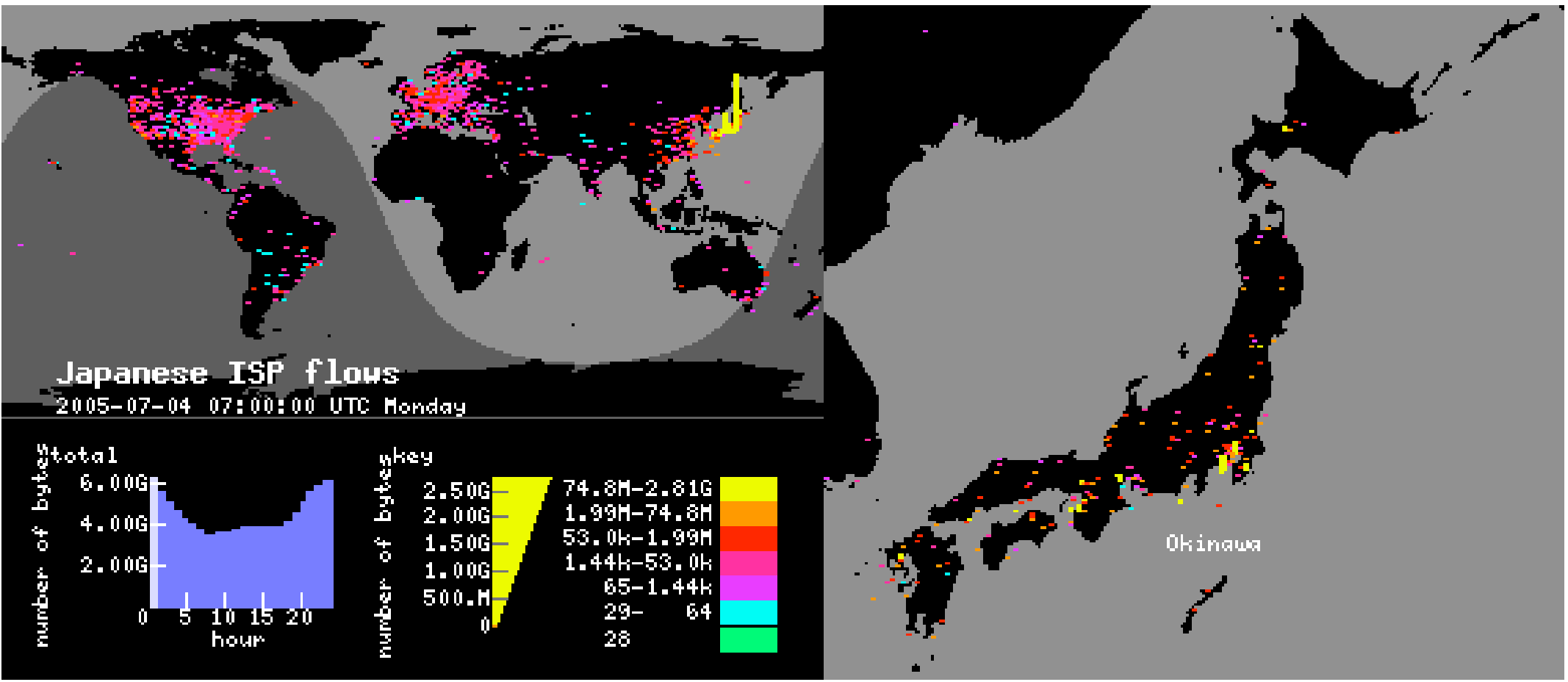
2. macroscopic topology measurement

Goal. (1) Enable the development of more realistic models of Internet topology based on actual data; (2) Compare performance issues in IPv4 vs. IPv6.

1. skitter -> scamper transition (in progress)
   (machines, dst list, storage mgt, analysis tools)
2. analyze collected scamper data (inc. validation)
3. develop methods of latency visualization
4. develop and test smarter probing algorithms

daunting place to do science

cuttlefish example: time-series geographic display

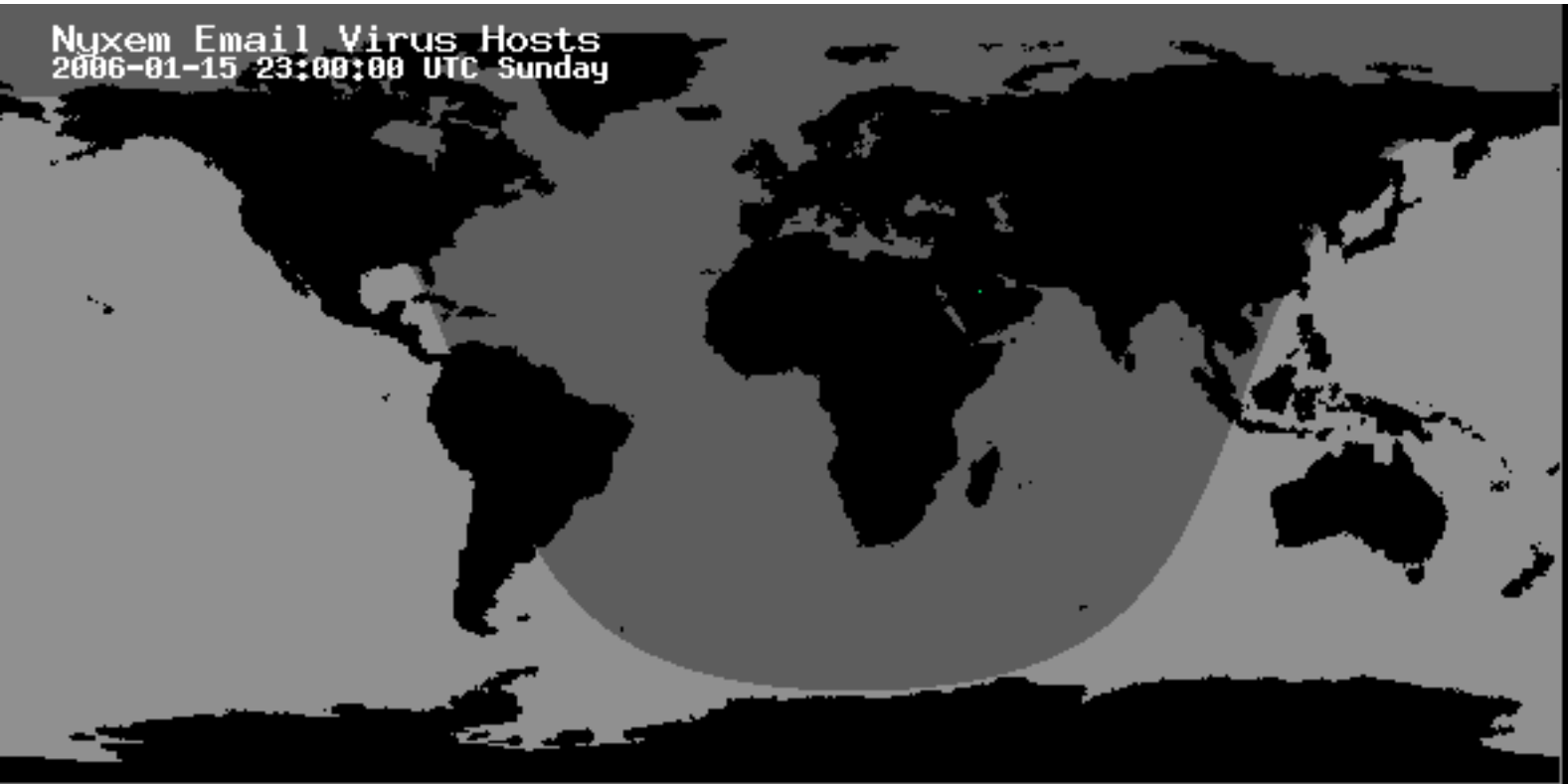


(general and configurable: can use with many kinds of data)

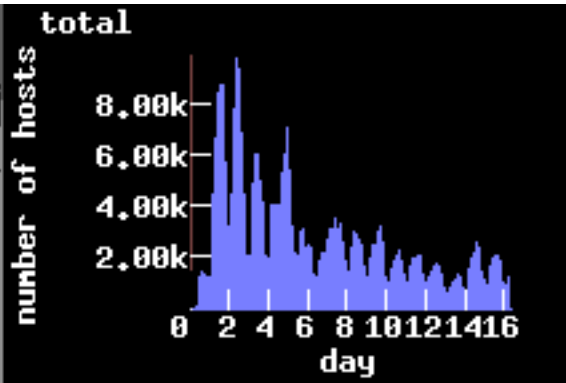
2. security analysis

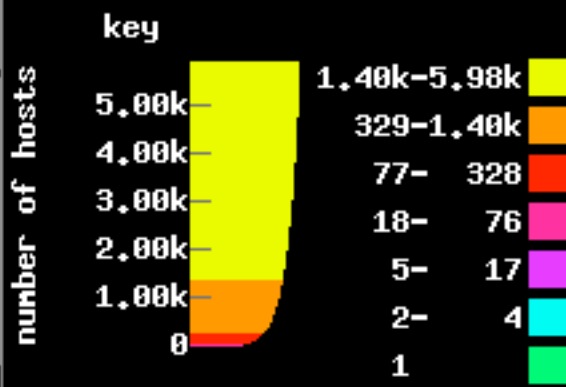
Goal. (1) analysis on network-based attacks (e.g. denial-of-service attacks), and measuring and analyzing the trends and impact that certain Internet worms and viruses have on the global network infrastructure.

http://www.caida.org/analysis/security/


Nyxem Email Virus Hosts
2006-01-15 23:00:00 UTC Sunday


Nyxem Email Virus Hosts
2006-01-15 23:00:00 UTC Sunday


total
number of hosts
8.00k
6.00k
4.00k
2.00k
0 2 4 6 8 10 12 14 16
day


key
number of hosts
5.00k
4.00k
3.00k
2.00k
1.00k
0
1.40k-5.98k
329-1.40k
77- 328
18- 76
5- 17
2- 4
1

3. DNS (domain name service) analysis

Goal. (1) Contribute to better understanding of current DNS problems, provide empirical data to test proposed solutions.

1. dsc software improvements and deployment
2. public listing of vulnerable name servers
3. compare DNSSEC usage in Asia, Europe, US, etc (also UDP vs TCP in face of DNSSEC, IDN, IPv6)
4. 48-hour root name server collections (OARC) for all interested root name servers
5. integrate WIDE data w/ CAIDA data infrastr.

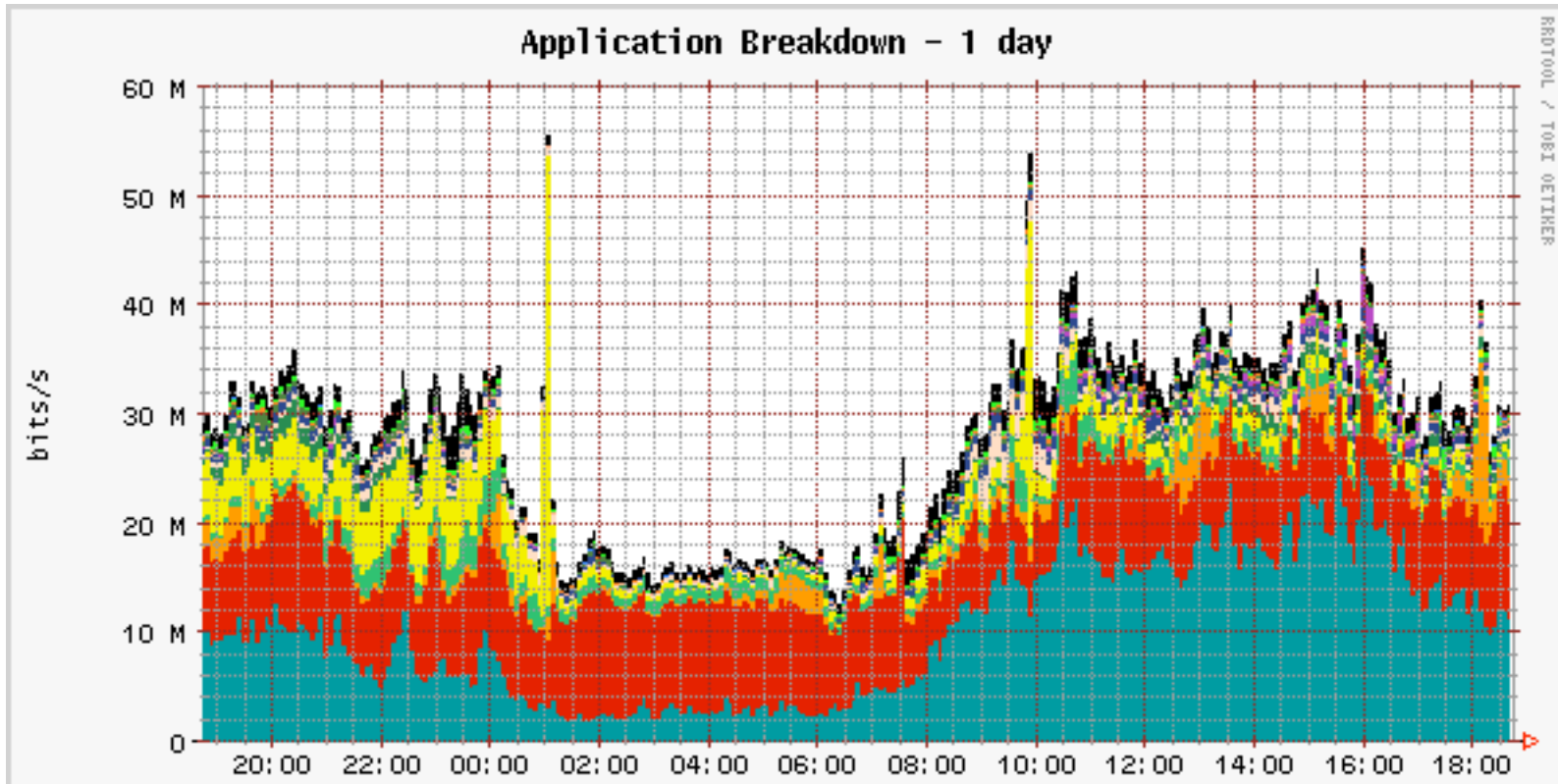
brings international data sharing issues to front...

4. "day-in-the-life of the Internet"

Goal. (1) Contribute to global understanding of Internet workloads, topology, performance, and routing, using cooperatively gathered empirical data. (2) improve science of and reputation of Internet.

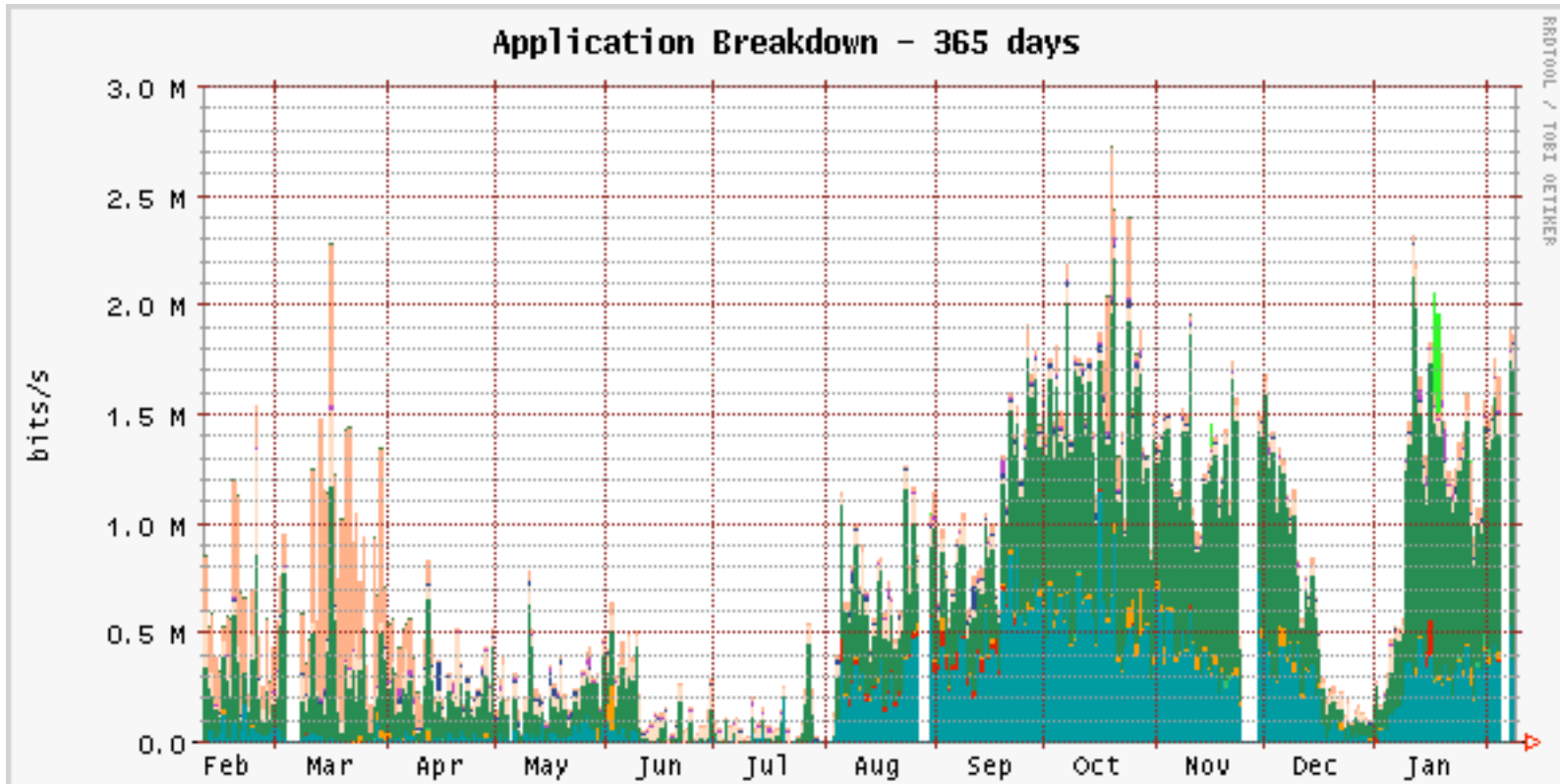
Motivation:

> *A reviewer of a draft of this report observed that this proposed framework: measure, develop theory, prototype new ideas -- looks a lot like Research 101. Why did this exploratory effort end up framing a research program along these lines? From the perspective of the outsiders, the insiders did not show that they had managed to execute the usual elements of a successful research program, so a back-to-basics message was fitting.*
>
> *"Looking Over the Fence: A Neighbor's View of Networking Research", Computer Science and Telecommunications Board, US Nat'l Academy of Science*
>
> *-- http://www.nap.edu/books/0309076137/html/13.html*



Application	Min	Avg	Max
HTTP (World Wide Web)	1.92 M	10.34 M	25.96 M
OTHER_TCP (Other TCP)	4.97 M	8.54 M	14.53 M
SSH (Secure Shell)	0.04 M	1.15 M	15.33 M
BITTORRENT (BitTorrent file sharing)	0.32 M	1.26 M	4.81 M
SMTP (Mail Forwarding)	0.14 M	2.63 M	41.29 M
OTHER_50 (Other IPSEC-ESP)	0.02 M	0.53 M	2.22 M
OTHER_UDP (Other UDP)	0.12 M	0.65 M	3.47 M
HTTPS (Secure Web)	0.04 M	0.62 M	1.77 M
FTP_DATA (FTP (data stream))	0.00 M	0.23 M	3.27 M
IRC (Internet Relay Chat)	0.02 M	0.08 M	0.25 M
SQUID (Squid Web Cache)	0.00 M	0.25 M	1.71 M
RTSP (Real Time Stream Control Protocol)	0.05 M	0.16 M	0.98 M
POP (Post Office Protocol (v2 & v3))	0.03 M	0.22 M	2.13 M
EDONKEY_TCP (eDonkey file sharing system (TCP))	0.00 M	0.03 M	0.32 M
FASTTRACK (FastTrack file sharing system)	0.00 M	0.00 M	0.24 M
other (other)	0.35 M	1.30 M	5.48 M

generated Wed Feb 08 18:44:21 2006 PST



Application	Min	Avg	Max
GNUTELLA (Gnutella file sharing system)	0.00 k	211.14 k	6181.97 k
NAPSTER_DATA (Napster MP3 Sharing (data))	0.00 k	23.85 k	1155.82 k
FASTTRACK (FastTrack file sharing system)	0.00 k	17.18 k	1628.47 k
IMESH_CTL (iMesh Sharing control)	0.00 k	0.99 k	1591.91 k
IMESH_DTA (iMesh Sharing data)	0.00 k	0.31 k	74.59 k
BITTORRENT (BitTorrent file sharing)	0.00 k	401.56 k	22355.29 k
REALAUDIO_UDP (RealAudio Player (UDP))	0.00 k	57.49 k	4460.78 k
QUAKE (Quake game)	0.00 k	8.05 k	5142.78 k
AOL (America Online)	0.00 k	6.03 k	2420.46 k
EDONKEY_TCP (eDonkey file sharing system (TCP))	0.01 k	82.55 k	2698.06 k
EDONKEY_UDP (eDonkey file sharing system (UDP))	0.00 k	0.07 k	33.38 k
SCOUR_EX (Scour Exchange music sharing system)	NaN k	NaN k	NaN k
DIRECT_CONNECT (Direct Connect file sharing)	0.00 k	200.13 k	3368.80 k

generated Wed Feb 08 18:49:24 2006 PST

4. “day-in-the-life of the Internet”

0. assumes that research community might be mature enough to handle this
1. several year project, policy and technical aspects
2. need researchers, sysadmins, lawyers, engineers, etc
3. e.g., compare file sharing usage in Asia, Europe, US, etc (also UDP vs TCP in face of DNSSEC, IDN, IPv6)
4. make it easy for people to contribute data
5. int’l frameworks a la http://www.predict.org

5. economic modeling

Goal: transparent analysis of cost-of-service-delivery of Internet access on realistic scale.

1. need access to economic data for backbones
2. currently proposing joint project with NLR (US)
3. need parallel efforts around world (JP-WIDE, EU-GEANT?)
4. 'cost-of-service-delivered' and other formal economic analyses

bad news: not just complex system, but economic one.
but we are obligated to formalize what we know.

6. scalable routing

Goal: mathematically rigorous next-generation routing protocols for realistic network technologies.

1. applying recent theoretical routing results to practical goal: fix Internet routing system
2. next step: extend theoretical results to more practically acceptable constraints
3. validate applicability against real topology data
4. build and evaluate model for Internet topology evolution, reflecting fundamental laws of evolution of large-scale networks

Workshop on Internet Topology, May 06
http://www.caida.org/outreach/isma/0605/

measurement accuracy is the only fail-safe means of distinguishing what is true from what one imagines, and even of defining what true means.

..this simple idea captures the essence of the physicist's mind and explains why they are always so obsessed with mathematics and numbers: through precision, one exposes falsehood.

a subtle but inevitable consequence of this attitude is that truth and measurement technology are inextricably linked.

-- robert b laughlin, a different universe,