Internet measurement: what have we learned?

kc claffy
kc@caida.org
19 may 06

outline

- motivation: ‘new Internet’ initiatives
- goal: highlight ten years of investigation
  - assess performance along (us)nsf criteria: (1) intellectual merit, (2) broader impact
- identify roots of limits to current progress
- consider implications for future of Internet measurement as well as network research and public policy

falling bits of sky

The Twenty Most Critical Internet Securit

Version 6.01 November 28,
Questions / comments
To link to the Top 20 List, use the SANS

-Jump To Index of Top 20 Vulnerabilities -----

oduction

SANS Top 20 Internet Security Vulnerabilities

r years ago, the SANS Institute and the National Infrastru
ter (NIPC) at the FBI released a document summarizing t
rnet Security Vulnerabilities. Thousands of organizations
anded Top-20 lists that followed one, two, and three year
r efforts so they could close the most dangerous holes fi
vices that led to worms like Blaster, Slammer, and Code
se lists.

SANS Top-20 2005 is a marked deviation f
itic
lic
ev
pr
ed
he
ch
ha
ce
ne
ea
er
do

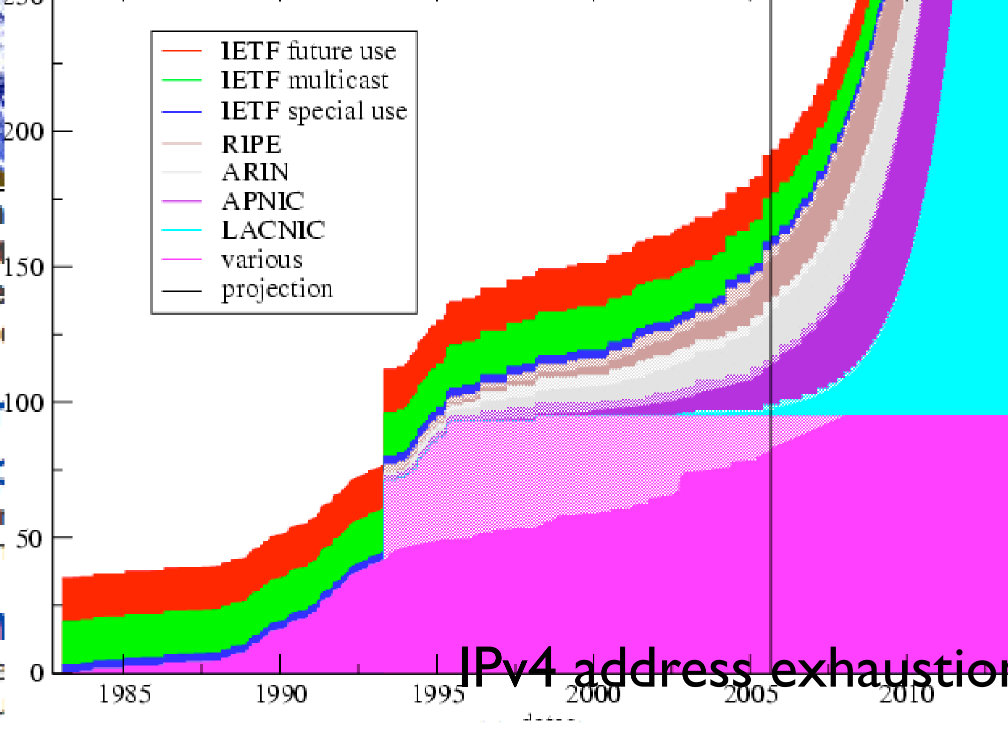
The In
securit
can he
and m

Sear
Resp
Opinio
Shouldr

High
The ope
that cou

The Dark Side of the Search Engine Business

Paid search is a booming business for Google, Yahoo and Microsoft, but there's a major downside for users. A new study by McAfee's SiteAdvisor



IPv4 address exhaustion

Home >> China

China adds top-level domain names

China's Ministry of Information Industry (MII) has ma
domain name system in accordance with Article 6 of
Regulations.

After the adjustment, ".MIL" will be added under the
"CN".

A new Internet domain name system will take effect

Under the new system, besides "CN", three Chinese
"NET" are temporarily set. It means Internet users d
servers under the management of the Internet Corporation for Assig

How the internet killed the phone busines

Almost-free internet phone calls herald the slow death of traditional telephony

THE term "disruptive technology" is popular, but is widely misused. It refers not simply to a clever new technology, but to one that undermines an existing technology—and which therefore makes life very difficult for the many businesses which ... the existing way of doing things. Twenty years ago, the personal computer was a classic example ... pushed aside an older mainframe-based style of computing ... eventually brought IBM, one of the world's mightiest ... party of sorts for another disruptive technology, "voice over

market, as the marginal price of making phone ... exorably downwards.

VOIP makes possible more than just lower ... ever. It also means that, provided you have a br... nection, you can choose from a number of prov... telephony and related add-on services, such ... conference calling or video. Many providers all... count to be associated with a traditional tel... or multiple numbers. So you can as... Francisco number, a New York number and a Lo... when your computer or VOIP phone—and then ... a local call by anyone in any of those cities. Furthermore, your phone (or computer) will ... you are in the world, as soon as it is plugged in

http://www.economist.com/ Sept 2005

International Telecommunication Union

Home : Office of the Secretary General : SPU

Building Partnerships for Progress

All match
More

Home: OECD > OECD ICCP Workshop: "The Future of the Internet", Paris, 8 March 2006

OECD ICCP Workshop: "The Future of the Internet", Paris, 8 March 2006

THE INTERNET.COM

Fight for Internet Freedom

THE LATEST....

Moby Speaks Out on Internet Freedom

At a press event in Washington today, Grammy-nominated musician Moby (along with Rep. Ed Markey of Mass.) introduced Artists and Musicians for Internet Freedom, an

IPv6

From Wikipedia, the free encyclopedia

Internet Protocol version 6 (IPv6) is a network layer standard used by electronic devices to exchange data across a packet-switched internetwork. It follows IPv4 as the second version of the Internet Protocol to be formally adopted for general use.

IPv6 is intended to provide more addresses for networked devices, allowing, for example, each cell phone and mobile electronic device to have its own address. IPv4 supports 4.3×10^{9} (4.3 billion) addresses, which is inadequate to give one (or more if they possess more than one device) to every living person. IPv6 supports 3.4×10^{38} addresses, or 5×10^{28}(50 octillion) for each of the roughly 6.5 billion people alive today.

Invented by Steve Deering and Craig Mudge at Xerox PARC, IPv6 was adopted by the Internet Engineering Task Force in 1994, when it was called "IP Next Generation" (IPng). (Incidentally, IPv5 was not a successor to IPv4, but an experimental flow-oriented streaming protocol intended to support video and audio.)

RKWORLD RESEARCH CENTER:

Convergence / VoIP

NetworkWorld.com > Convergence / VoIP >

What IMS promises enterprises and carriers

Internet Protocol Multimedia Subsystem called key to converged, expanded services.

By Stephen Lawson, IDG News Service, 09/2

The latest buzzword in telecom isn't the name of a box, an application or a service. Instead, IMS is a way of organizing all those elements and more.

sundry "solutions"

The Future of the Internet

In a decade, the Net will dig deeper into our lives.

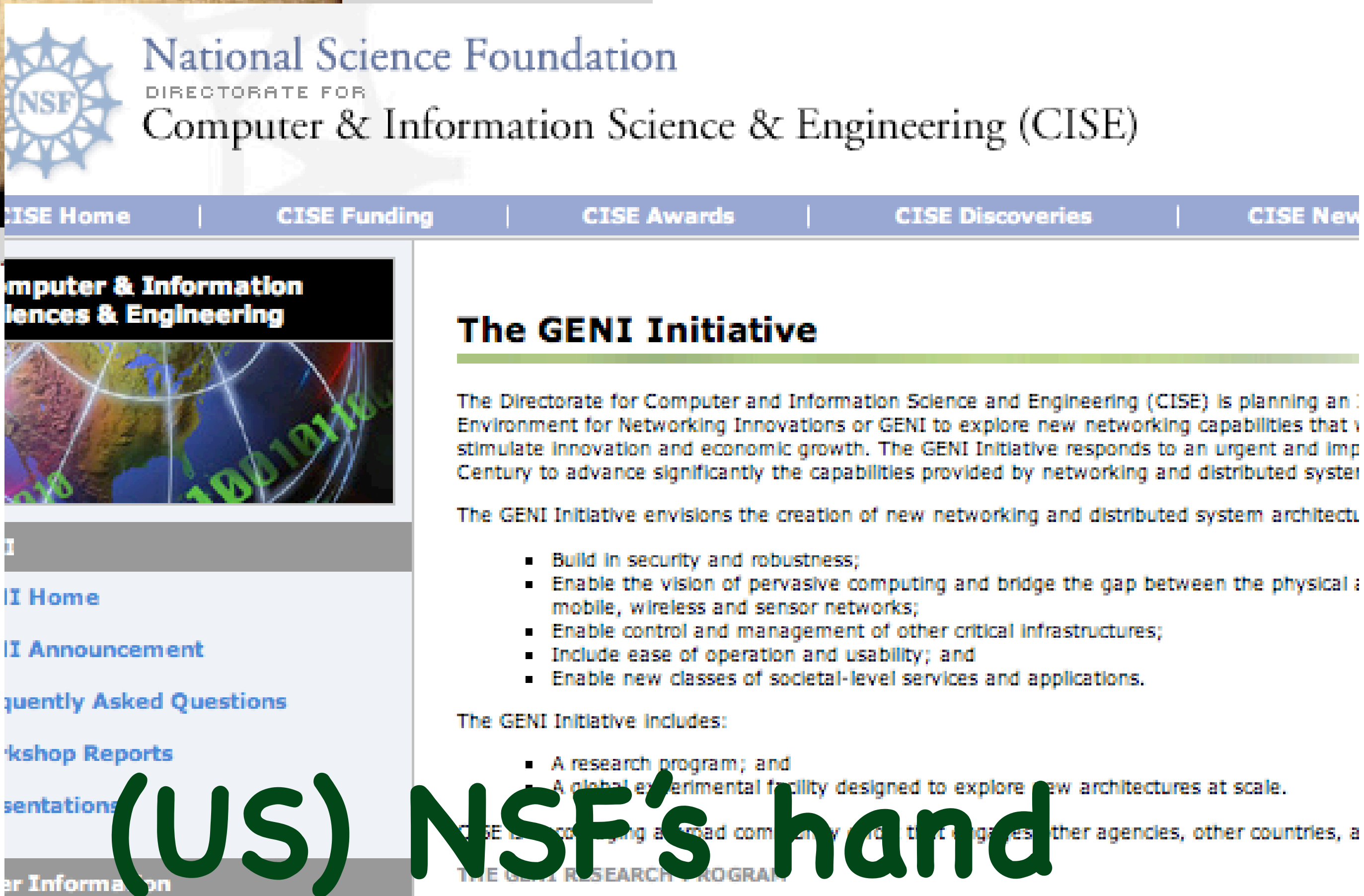
April 10, 2006 Issue

Credit: Dave Cutler

http://www.redherring.com

"While the business case for the carriers may be disappearing, a host of new business and investment opportunities is being created with far greater economic wealth creation," Mr. Arnaud writes in his blog. "Our biggest concern is that governments will be distracted by the complaints of the old industry such as carriers and penalize the new economy industries of the Internet."

"We don't presently have a roadmap of where we are trying to go with the Internet," says MIT's Mr. Clark. Instead of worrying about backward compatibility and migration issues, the focus has shifted to "where we would like to be in 10 to 15 years," he explains. "If the story is compelling enough, people will figure out how to get there."



National Science Foundation
DIRECTORATE FOR
Computer & Information Science & Engineering (CISE)

The GENI Initiative

The Directorate for Computer and Information Science and Engineering (CISE) is planning an
Environment for Networking Innovations or GENI to explore new networking capabilities that
stimulate innovation and economic growth. The GENI Initiative responds to an urgent and imp
Century to advance significantly the capabilities provided by networking and distributed syste

The GENI Initiative envisions the creation of new networking and distributed system architectu

- Build in security and robustness;
- Enable the vision of pervasive computing and bridge the gap between the physical a mobile, wireless and sensor networks;
- Enable control and management of other critical infrastructures;
- Include ease of operation and usability; and
- Enable new classes of societal-level services and applications.

The GENI Initiative includes:

- A research program; and
- A global experimental facility designed to explore new architectures at scale.

CISE is [illegible] a broad community [illegible] that engages other agencies, other countries, a

THE GENI RESEARCH PROGRAM

(US) NSF's hand

e.g. NSF’s GENI initiative

- US NSF responding to network research community frustration
  - difficulty with technology transfer, not to mention science
  - persistent problems leaking into unready world
- attempt to redesign components ‘in the light’
- what did we learn from measuring this one?

scope of field

- workload
- topology
- routing
- performance
- security
- geolocation

also:
standards,
software,
storage,
statistics.
and recently,
lawyers.

intellectual achievements

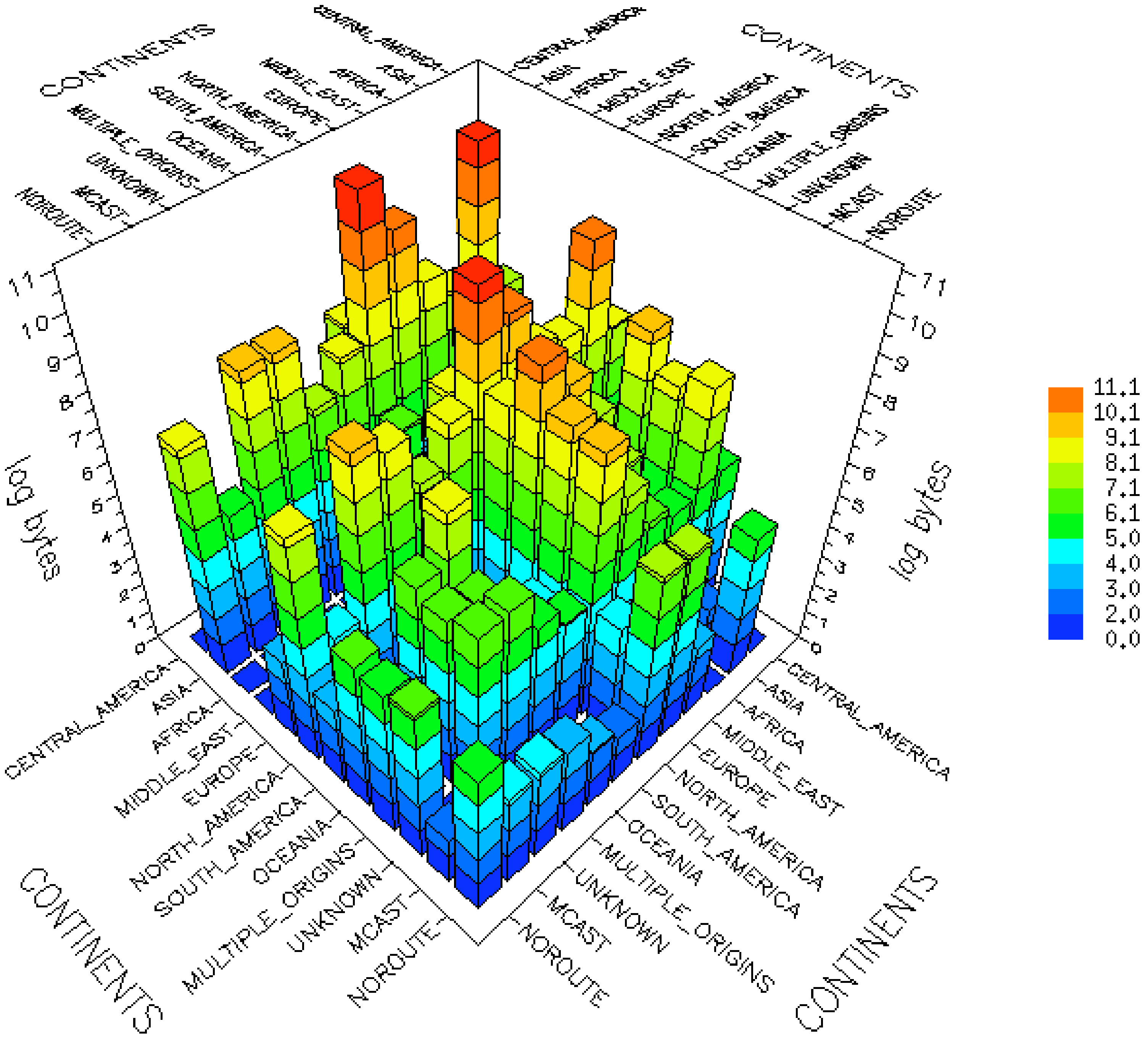
workload characterization & modeling

- traffic matrix inference (on small scale..we think)
- cross-section of core (failure, but lesson)
- self-sim/long-range dependence (on LAN networks)
- source-level (web object) models for LRD traffic
- intelligent sampling & anonymization methods

none generally used by vendors

intellectual achievements

traffic matrix visualization example



intellectual achievements

workload characterization & modeling

- flow menagerie (traffic engineering challenge)
- relentless growth in p2p (economic challenge)
- relentless growth in spam
- relentless growth in worms, viruses (recently a data src)
- critical infrastructure (dns roots) sees much (up to 80% of traffic) pollution

people use connectivity once there

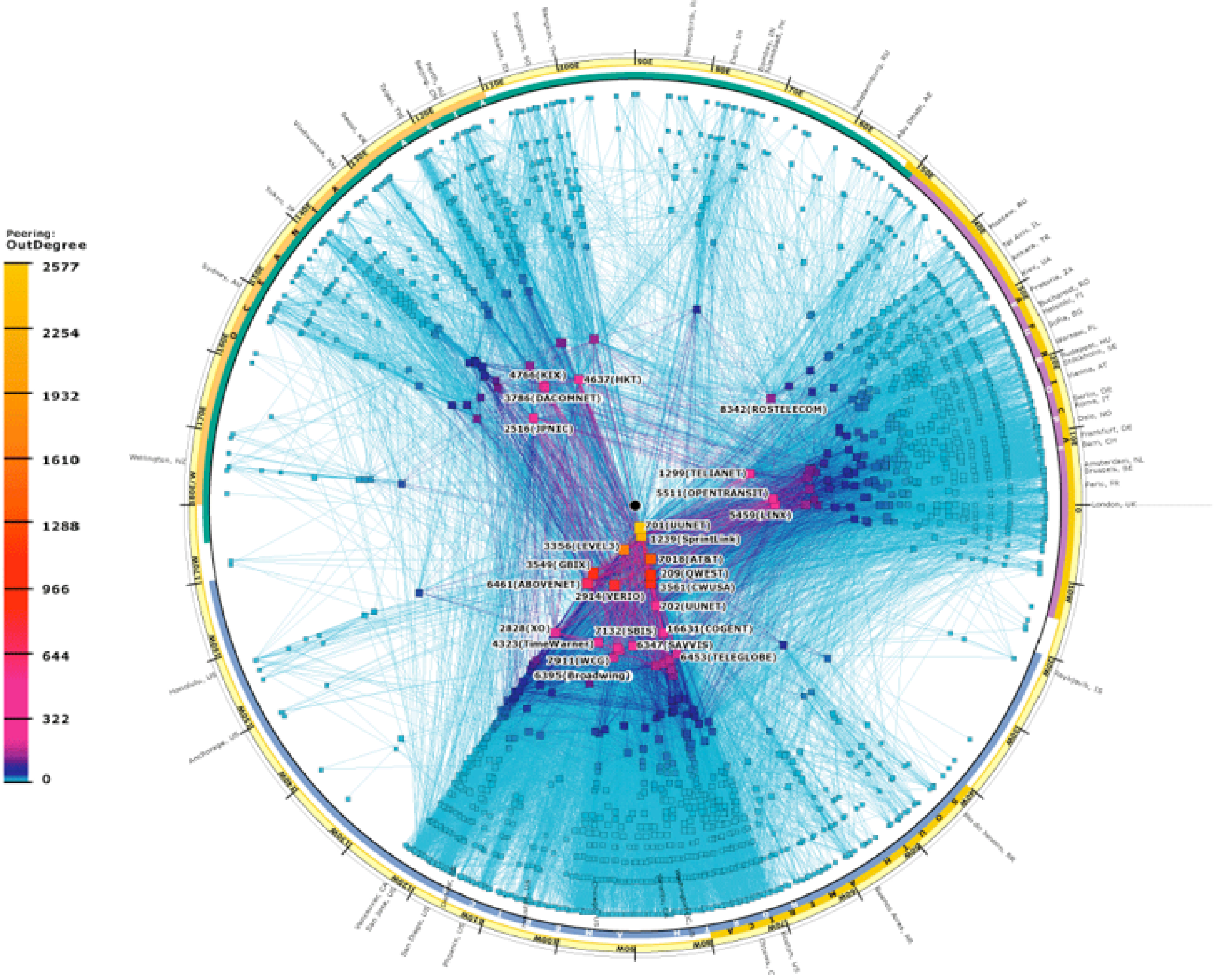
topology structure and dynamics

- not just random (see google) -- degree variability higher than expected.
- power law distributions (AS, router degree), or not.
- degree distribution doesn’t fully describe a graph, correlations not understood (forced vs natural)
- small distance distributions implies current (& proposed) routing architectures inherently poor fit

top-down vs bottom-up tension:

fit data or explain phenomena (former is easier)

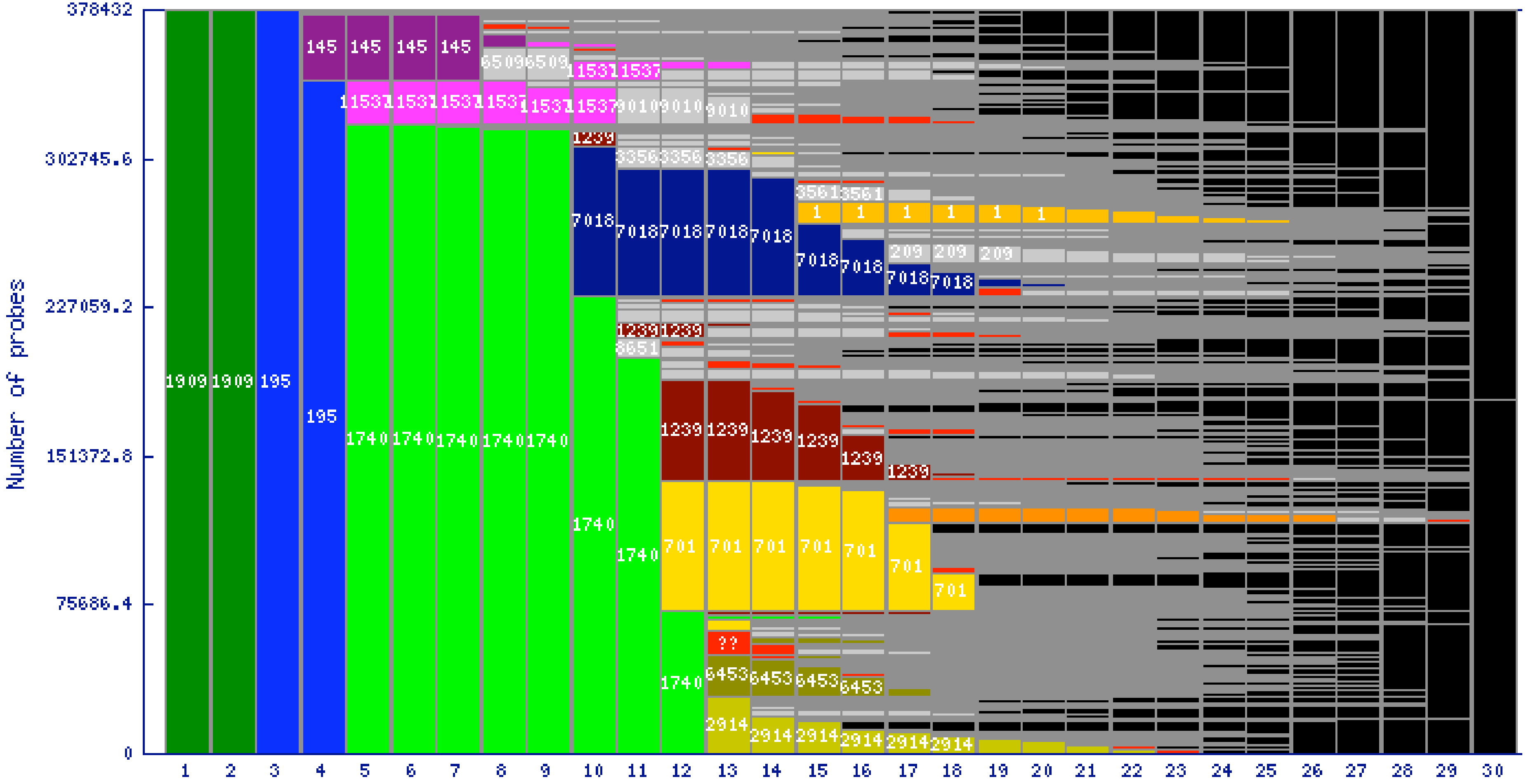
intellectual achievements



AS topology structure

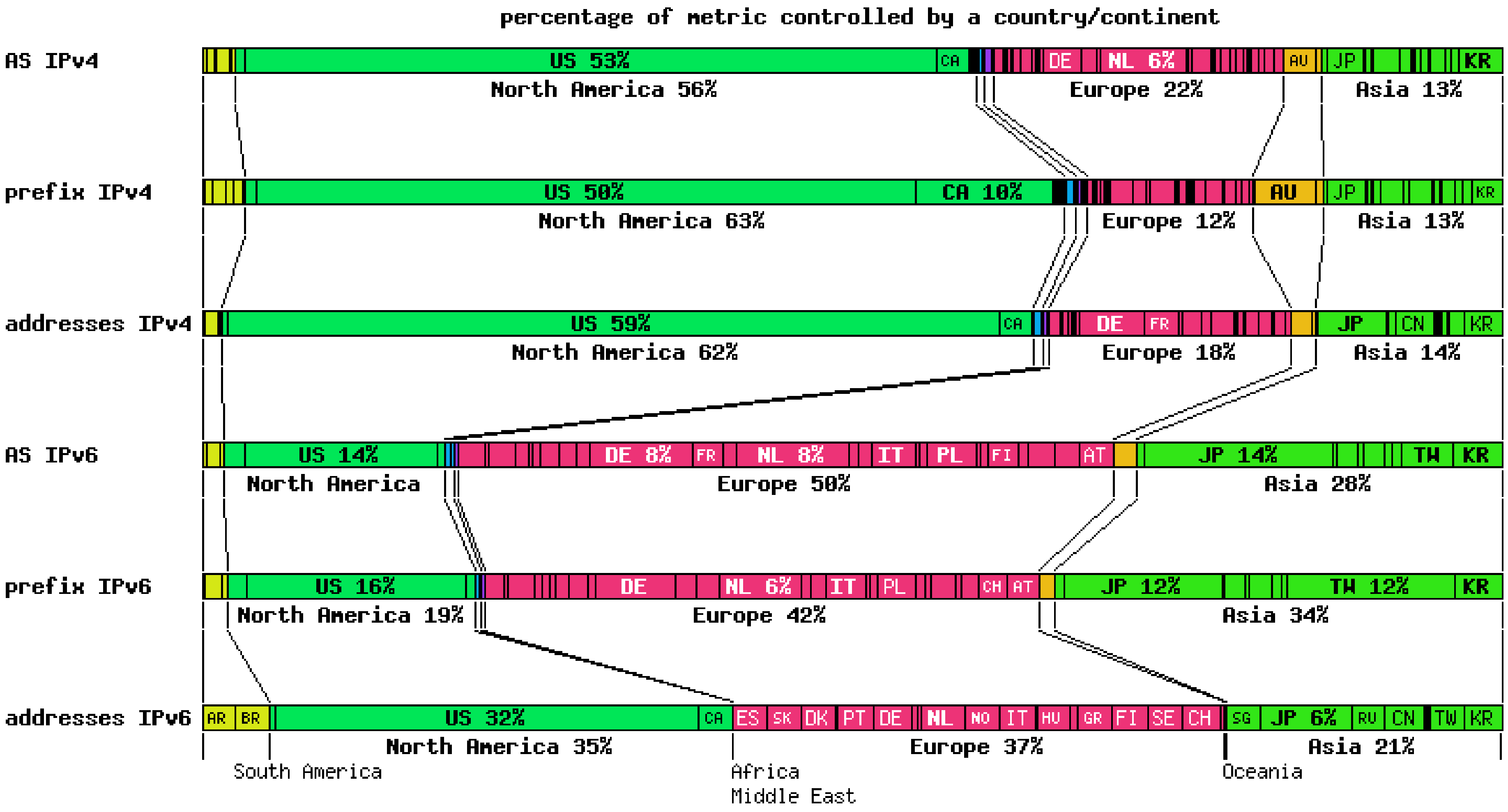
intellectual achievements

topology structure and dynamics



- AS dispersion from single source/many dests

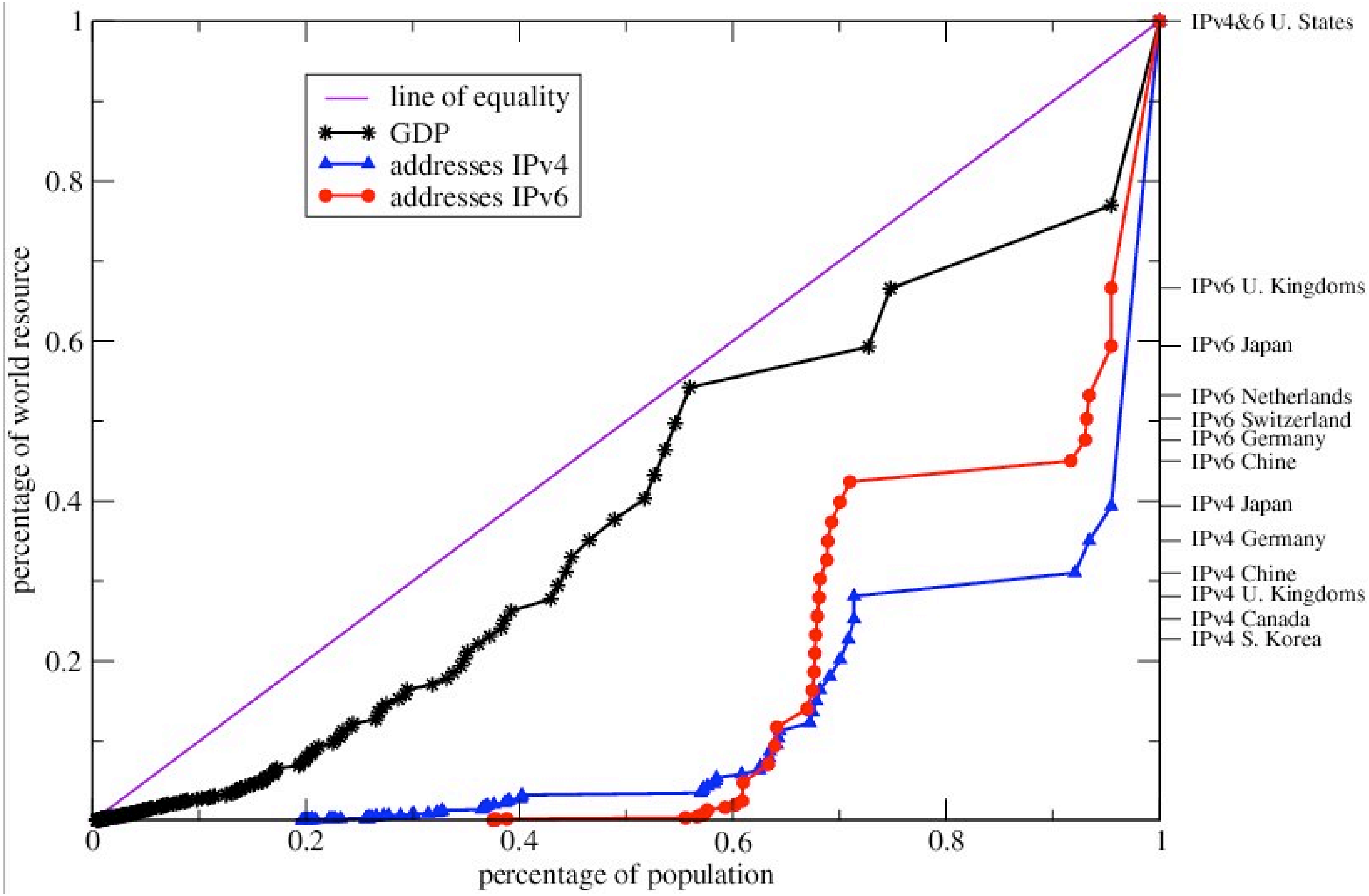
‘topology’ vs geography



- allocated AS and IP address space

intellectual achievements

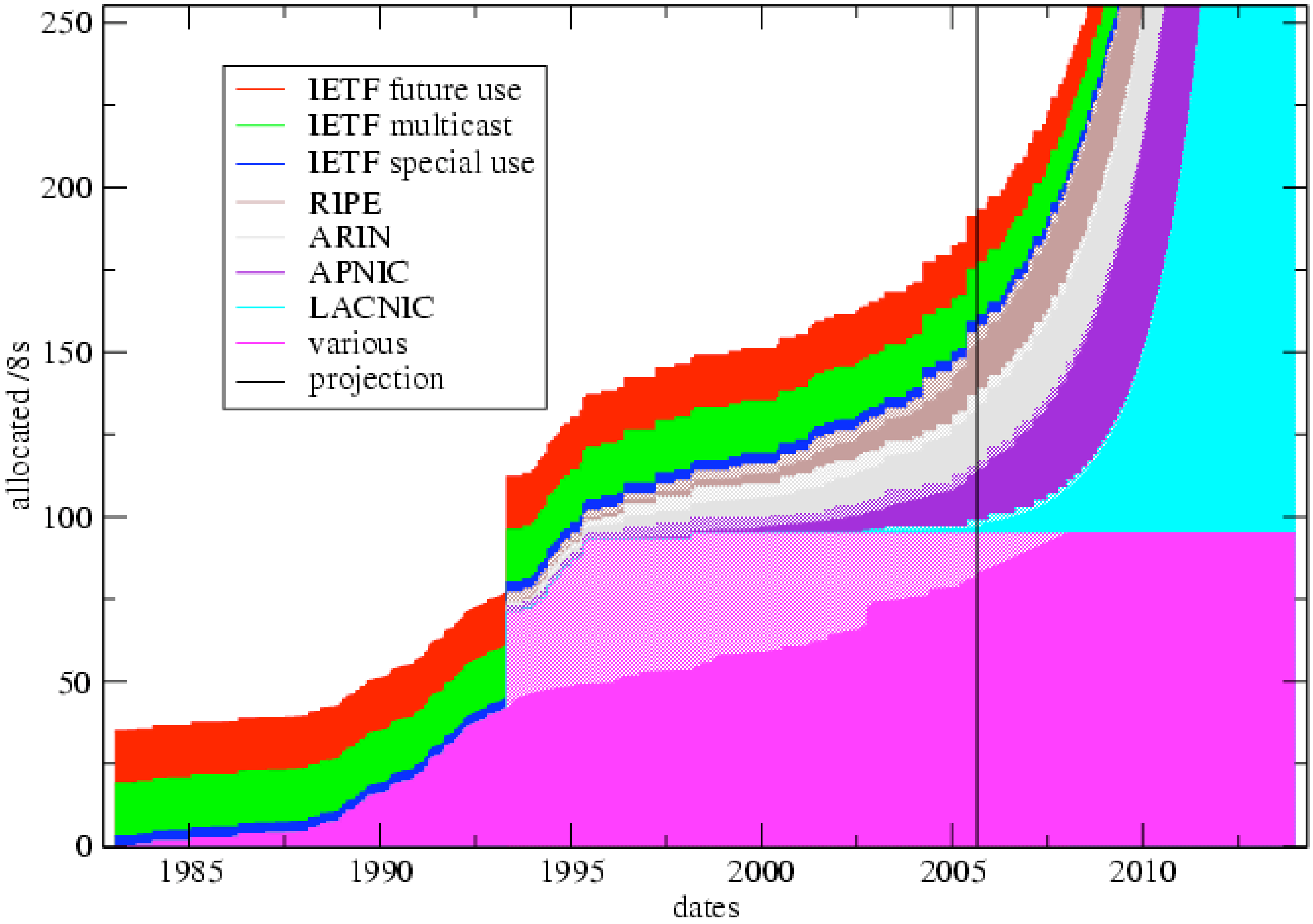
address resource distribution



Lorenz curve of inequality

IPv4 allocated /8s (first)

RIR whois dumps and IANA table of top-level /8 allocations



routing

- among hot topics in global Internet neurology: AS relationship inference, security, anomaly detection, configuration engineering, intelligent routing, sensor, adhoc, delay-tolerant, policy framework. validation hard.
- discovery: persistent oscillations observed, but if we follow certain simple rules, we can achieve stability. but no way to enforce simple rules.
  - BGP has inherently non-deterministic features (MEDs)
- discovery: observed evolving topology diverging from current (and proposed) routing system.

recognized need for new routing architecture

(and yet noone wants to bring it up)

performance

- distance-estimation methods, limited
- ECN, RED, CBQ: developed, not deployed
- bandwidth estimation: failed at per-link, can do limited per-path, not deployed
- systems integration complexity hinders validation
- unvalidated commercial ‘achievements’, e.g., keynote, internetweather, akamai, corporate SLAs

daunting place to do science

(don’t know congestion locations, lengths, or causes)

intellectual achievements

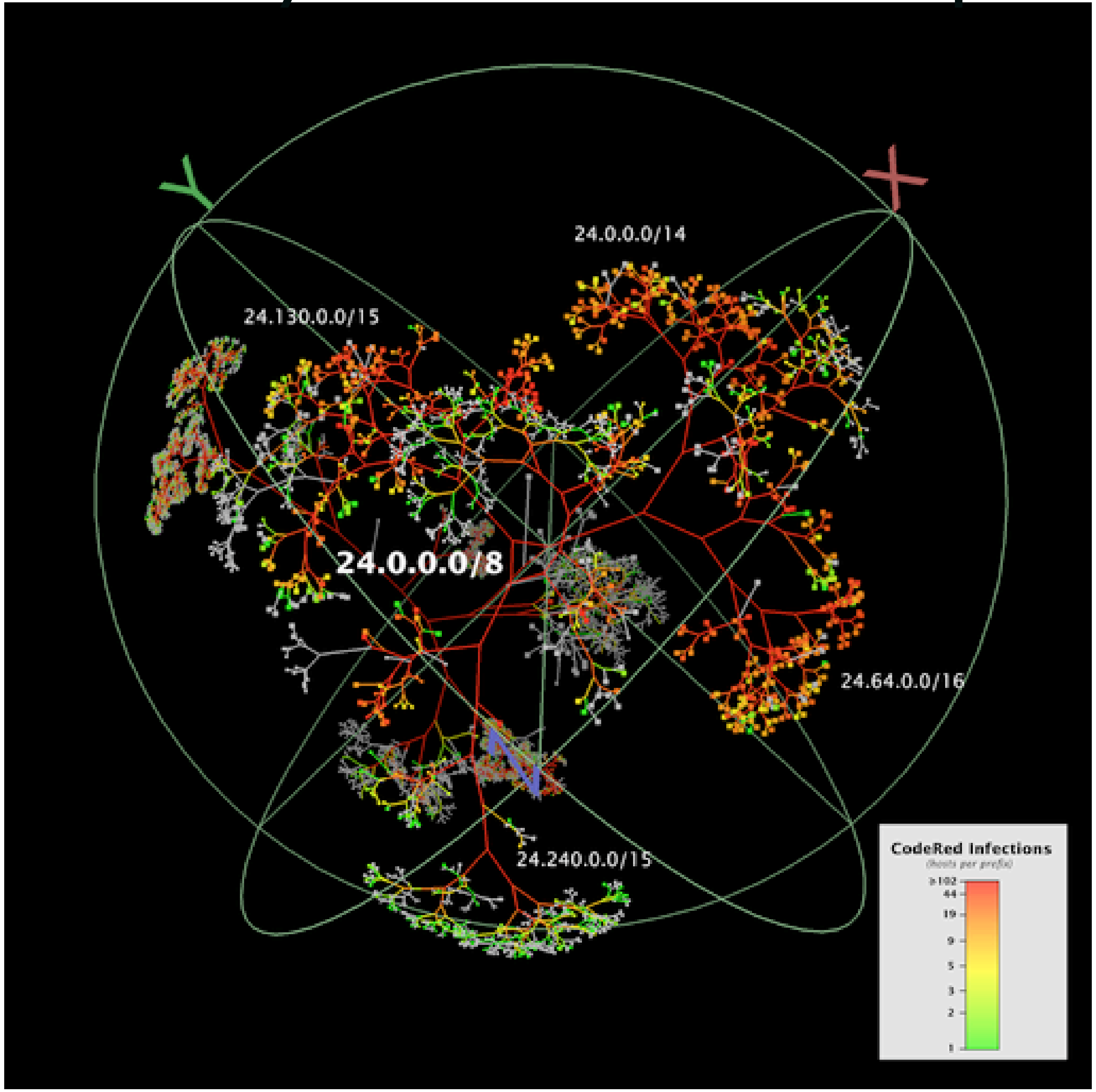
security

- detection & mitigation of specific (similar) threats
- worm propagation models, intrusion detection tools, even traceback startups
- discovery: patching model a failure
- discovery: monoculture a failure
- discovery: can’t quarantine networks fast enough
- discovery: correlated attacks (e.g., botnets) prevalent
- discovery: little ingress filtering; open (vulnerable) DNS resolvers

hard to measure progress of a given innovation,
scope of attacks & number of vulnerabilities
guarantees thriving industry w or w/o science

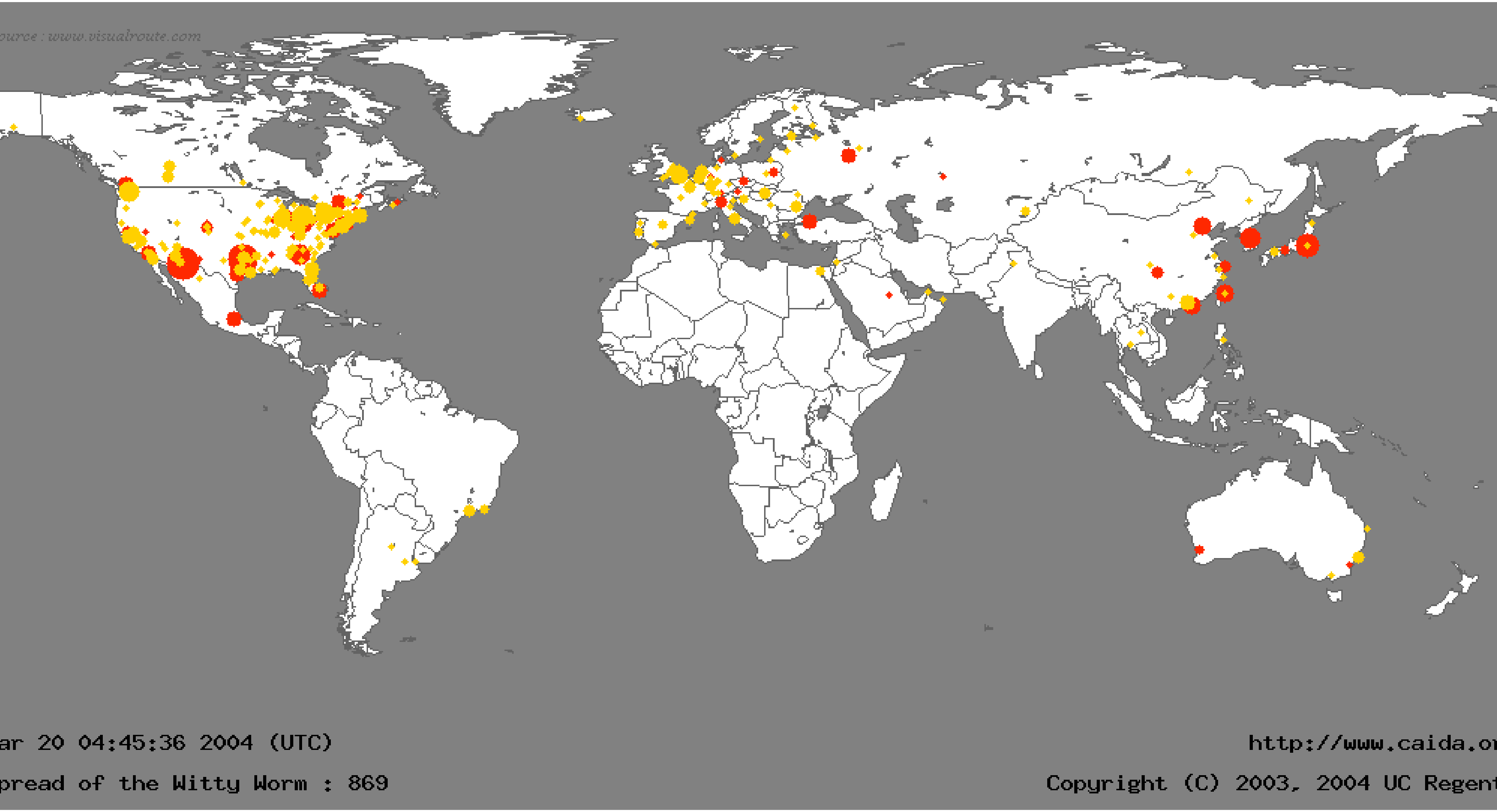
intellectual achievements

security: visualization example

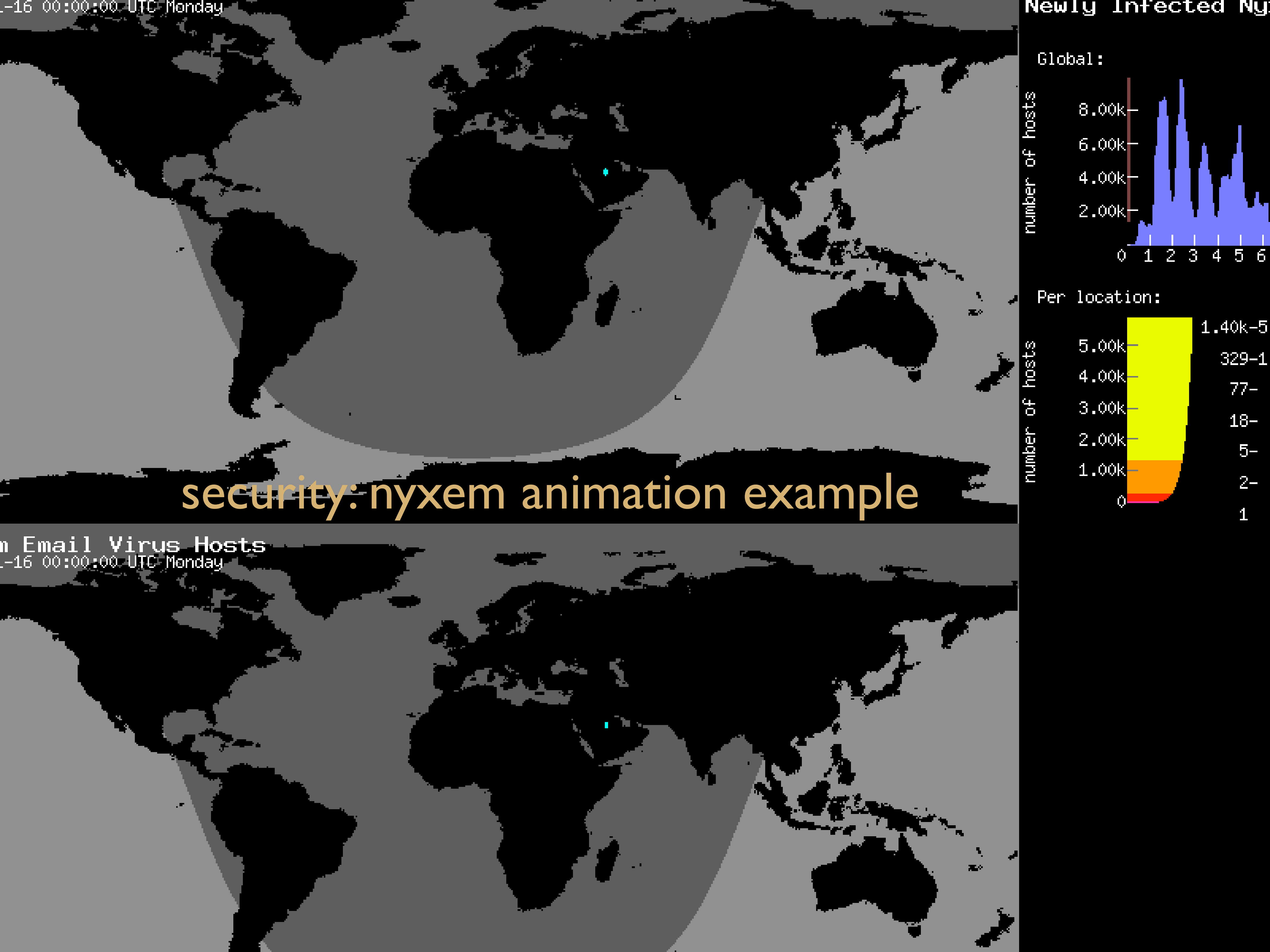


- prefix colored by number of infected hosts

security: animation example


ource : www.visualroute.com
ar 20 04:45:36 2004 (UTC)
http://www.caida.or
pread of the Witty Worm : 869
Copyright (C) 2003, 2004 UC Regent

security: nyxem animation example



geolocation

- identifying location of IP address
  - mapping, marketing, localization, server selection, law enforcement
- using dns, traceroute, whois, RTT, triangulation, metro location of an IP address generally possible. but kludgy.
- guaranteed validation requires human

continued R&D of heuristics and databases but not considered science so funding-starved

notable achievements under circumstances

for U.S. inter-domain internet science, the crash happened in 1994 when the nsfnet retired...

- can't figure out where an IP address is
- can't measure topology effectively in either direction, at any layer
- can't track propagation of a routing update across the Internet.
- can't get router to give you all available routes, just best routes
- can't get precise one-way delay from two places on the Internet
- can't get an hour of packets from the core
- can't get accurate flow counts from the core
- can't get anything from the core [we used to have anonymized traces]
- can't get topology of core
- can't get accurate bandwidth or capacity info
  not even along a path, much less per link
- can't trust whois registry data
- no general tool for `what's causing my problem now?'
- privacy/legal issues deter research (& was hard in enlight'd monarchy)

science abysmal, discouraging to remaining academics

by other measures it looks splendid

citations of measurement papers healthy:

“for 10 years, Internet measurement papers have been top 20 most-cited citeseer papers”

new conferences: IMW->IMC, PAM

standards of science not so healthy.

haven’t cultivated measurement culture

just starting to learn that tools from other disciplines sometimes work better than our own.

”the insiders did not show that they had managed to execute the usual elements of a successful research program...This report challenges the research community to develop the means to capture a day in the life of the Internet to provide such information.” -- Looking Over the Fence, National Academies, 2001

jarring observation from history of science

The modern field of elementary particle physics depended crucially on the establishment of a huge volume of data gathered mainly in the period 1945-65. Only then was it possible for the synthesis of the Standard Model to take place, 1967-74.

-- Peter Galison, Professor of History of Science and Physics, Harvard

*(unfortunately, we're not doing research,
we're building critical infrastructure.
and it's riddled with structural problems.)*

broader impact

- what has happened to the Internet since the NSF transitioned it to the private sector “(commercialization and privatization”)?
- what false assumptions do we carry?
- for remaining problems, what is progress blocked on?
- how can we move forward?

broader impact

16 operational internet problems

- security
- authentication
- spam
- scalable configuration management
- robust scalability of routing system
- compromise of e2e principle
- dumb network
- measurement
- patch management
- “normal accidents”
- growth trends in traffic and user expectations
- time management and prioritization of tasks
- stewardship vs governance
- intellectual property and digital rights
- interdomain qos/emergency services
- inter-provider vendor/business coordination

persistently unsolved problems for 10+ years

why we’re not making progress

- top unsolved problems in internet operations and engineering are rooted in **economics, ownership, and trust (EOT).**
- even the most theoretical computer scientists are convinced.

does not mean there are not useful technical problems to work on. but there will no technical solutions that don’t solve the EOT problems.

warning: there’s a problem we left out

- the economics one runs deep
- best available data suggests that moving IP packets around is not even a for-profit enterprise. not just bernie ebbers factor.
- like most large scale transport networks (!)
- even harder to get sound economic data

noone tasked with thinking about the 25-year internet provisioning problem.

historical context

1966: Larry Roberts, "Towards a Cooperative Network of Time-Shared Computers" (first ARPANET plan)

(*we are still using the same stuff*)

1969: ARPANET commissioned by DoD for research

1977: Kleinrock's paper "Hierarchical Routing for large networks; performance evaluation and optimization"

(we are still using the same stuff)

1980: ARPANET grinds to complete halt due to (statusmsg) virus

1986: NSFNET backbone, 56Kbps. NSF-funded regionals.
IETF, IRTF. MX records (NAT for mail)

1991: CIX, NSFNET upgrades to T3, allows .com. web. PGP.

1995: under pressure from USG, NSF transitions backbone to competitive market. no consideration of economics or security. kc proposes caida.org

2005: *The Economist*'s cover story: "*How the Internet killed the phone business*" (September) broader impact

how unregulated players survive operating in an inherently non-profit industry

- hide the fact that you lose money by using non-IP revenue to subsidize developing IP habit. e.g., voice
- file bankruptcy every few years, includes billions spent on lobbying to keep incumbents in power rather than analyzing the macroproblem
- lie to the markets to get capital, confuse markets for a decade. or two. count on folks not reading history.
- long term: complex vertical integration (bad for security), infrastructure control (bad for freedom)
- don’t let anyone look at the data that would facilitate analysis of provisioning models for this commodity. don’t promote research & analysis.

what have we learned?

- most important thing we’ve learn so far: society has decided IP is like water.

- strong implications for an industry structuring itself to sell wine. but that’s what the data shows.

- when you want to move water, you care about 4 things: safe, scalable, sustainable, stewardship.

the 4 S’s

- safety: is the data toxic upon arrival?
- scalable: can we route/name/address earth’s needs?
- sustainable: is it economically viable?
- stewardship: will the provisioning and legal frameworks we choose leave our children -- and democracies -- better or worse off?

none of these are purely technical issues, but they all require deep technical (among other) understanding to get right.

and they’re all connected.

how have we done?

- how safe is the Internet?
  - data doesn’t look good
- how scalable is the Internet?
  - data doesn’t look good
- how sustainable is the Internet?
  - data doesn’t look good
- how did we do on stewardship?
  - data doesn’t look good

failure (to measure progress) on 4S’s poses risks to economics and

- that we won’t learn from our own history. e.g.,not only don’t we understand the economics, but we don’t understand that we don’t understand the economics, and thus must set policy based on unvalidated assumptions
- that we will design another architecture with no actual plan for economic sustainability (much less incenting further innovation in a competitive market!)
- that other forces will “code” innovation into the architecture (free markets vs free speech)

there is good news

- we made something so great, everyone wants it.
- in fact many of us want it more than once! (um..)
- the current industry is a historical artifact of technical and (science & regulatory) policy ‘innovations’ in the 60s, 70s, 80s, 90s, and 00s
- people are starting to study interplay, but they’re undercapitalized
- in the meantime, it became global critical infrastructure. oops.

broader impact

“science of the Internet”

The wonderful thing about science is that eventually nature tells you when you are fooling yourself. real objects can be measured again and measured by somebody else -- false signals will eventually be weeded out.

Robert Kirshner, The Extravagant Universe

but if what you need to measure is economics..

Knowing what to measure and how to measure it makes a complicated world less so. if you learn how to look at data the right way, you can explain riddles that otherwise might have seemed impossible.

Steven Levitt, Freakonomics

cataloguing lessons

- although the Internet has over-achieved on plenty, it has underachieved on: security, scalability, sustainability, and stewardship. substantial oversights.
- our ability to measure is surprisingly abysmal, although policy history explains
- cooperative, data-sharing approaches key to moving forward

we have learned more from our failures than from our successes...

implications for science policy

- confront data acquisition problem head-on
  - muni networks will help, still need lawyers
  - access to economic data is fundamental
  - US agencies, registries, OECD asking Internet researchers for help, FTC may follow
- bring standards of rigor to network science, promote interdisciplinary conversations, approaches
- top-down and bottom-up approaches, in collaboration with ISPs and governments

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,