Mapping the Internet to Assist Cyber-Defense

John Heidemann
joint work with Genevieve Bartlett, Xue Cai, Xun Fan,
Ramesh Govindan (co-PI), Christos Papadopoulos (co-PI),
Lin Quan, Yuri Pradkin
USC/ISI, USC/CS, and Colorado State U.

26 May 2010

Cyber-crime: what do we know?

what appears in the news…

"Digital Fears Emerge After Data Siege in Estonia", 29 May 2007

"Tackling cybercrime: guidance on sharing Internet data", 2 April 2008

"Surging Losses, but Few Victims in Data Breaches", 27 Sept. 2006

The New York Times

"The Evolution of CyberCrime, Inc.", 4 April 2008

who comes to us…

criminal reports

336,655 complaints in 2009

break-in resolutions and recovery

$559.7M loss reported to law enforcement in 2009

but not:

where are the attackers *from*?

where is risk *disproportionate*?

how to *evaluate risk* to people and infrastructure?

Need Internet Data to Improve Security

many network security questions…

- how **robust** is…
  - the Internet routing topology?
  - cloud computing?
  - models of topology, traffic, …?
- **trends** and correlations in…
  - {network location, country, provider, users…}
  - to {compromised hosts, spam generation, botnet C&C, traffic generation, service use…}
- new technology **deployment**…
  - firewalls? content filters?
  - IPv6 and new protocols? new applications?

what if an Internet exchange point was taken out?

- anecdotes are useful starters…

ex: Nov '08: take out a bad ISP (McColo)

how much did spam change?
how disproportionate was McColo's spam?
compare McColo to daily spam ebb&flow?

answers must combine spam + population

how many people still run vulnerable IIS versions?

One Insight: Active Probing

- we can *directly measure* the whole Internet
  - computers and networks are fast
  - there are "only" 4 billion IPv4 addresses
  - => a *census:* full enumeration in ~2 months
    estimate Internet size *and* new, general approach
- imperfect! (ex.: firewalls reduce coverage 40%)
- but we can *estimate and correct for error*

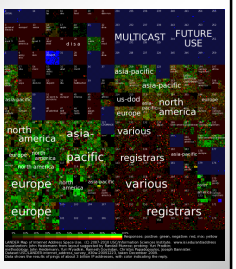
calibrated measurement provides real data

Our Approach

provide **calibrated data** and **browsable Internet maps** that are

- **updated**
  - continuous probing and regular map updates
- **edge-conscious**
  - end hosts and servers (not just routers and links)
- **annotated**
  - latencies, services, owners (not just connectivity)

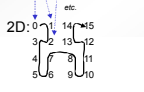
to improv our understanding and security of the net

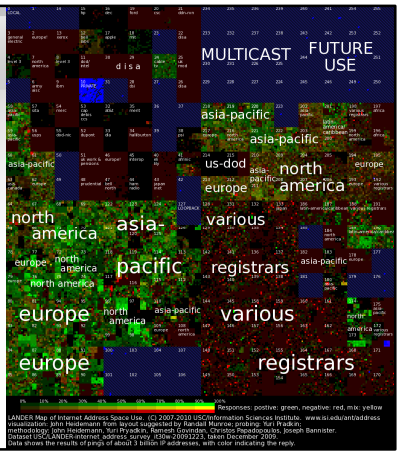


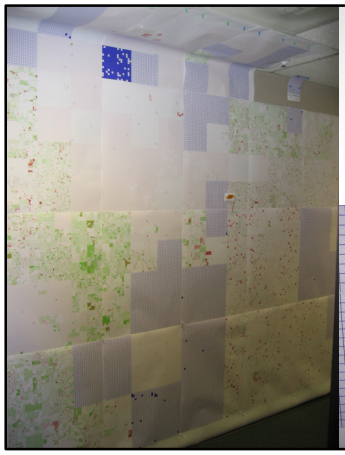
one map: pings to 3 billion addresses collected over 2 months

The Internet

- average each /16
  - each pixel: 65k addresses
  - represents all 2^{32} addrs
- brightness: responsiveness
- green/red-ness: degree of positive vs. negative replies
- blue: areas not probed
- layout: Hilbert Curve







The Whole Internet

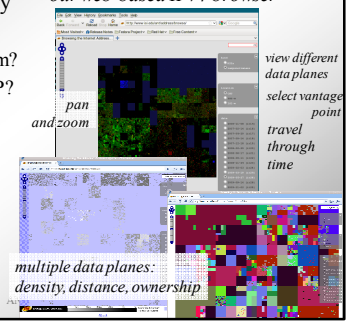
- here, 1 pixel is 1 address
- 9x9' at 600dpi
- green: positive, red: negative; white: no resp.

A Browsable Map for Network Situational Awareness

- needs of cyber-security first responders:
  - where is an attack from?
  - what jurisdiction? ISP?
  - type of host: home computer vs. server

=> network situational awareness

our web-based IPv4 browser



pan and zoom

view different data planes

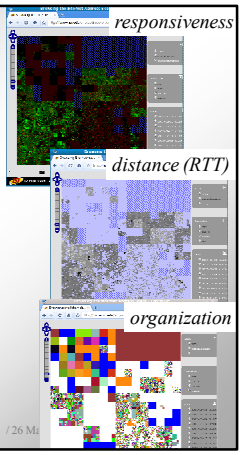
select vantage point

travel through time

multiple data planes: density, distance, ownership

Benefits of Better Information

- **new measurement tools**
  - run on the Internet
  - or *your* network
- **annotated maps**
  - data from running the tools
- studies **quantify accuracy**
  - answers *with* error bars
- **visualization and understanding** the results
  - browse, query, or reuse



responsiveness

distance (RTT)

organization

Additional Information Sources

- cyber-crime reports
  - important, but reactive—only part of picture
- end-user telephone surveys
  - important, but limited by general knowledge
  - "Q: Is your computer a zombie in a botnet?" A: "No, I am a mage in WoW, not the undead."
- other Internet measurement studies
  - important, but focused on the core, not the edge
  - (great for ISPs, not sufficient for cyber-defense)

Can Topology and Data Inform Your Work?

http://www.isi.edu/ant/address/browse/

browse today!

- browse our data on the web
- do our results apply to your analysis?
  - papers at *http://www.isi.edu/ant/pubs/*
- does our data fit your simulations and models?
  - data is free (gratis): *http://www.isi.edu/ant/traces/*
  - approval at *http://predict.org*
- can assist in focused scanning?
  - one-time or repeated views
  - talk to us: johnh@isi.edu