USC/LANDER Passive and Active Data Collection

John Heidemann
joint work with Genevieve Bartlett, Xue Cai,
Maureen Dougherty, Ramesh Govindan, Christos Papadopoulos (co-PIs),
Lin Quan, Yuri Pradkin
USC/ISI, USC/ITS, CSU

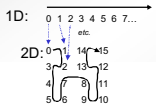
12 February 2009

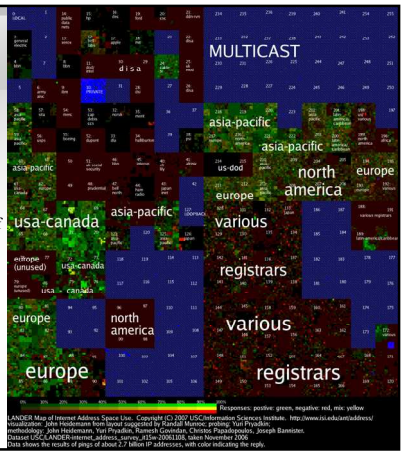
LANDER in One Slide

- http://www.isi.edu/ant/lander
- passive data collection (ongoing)
  - snooping all packet headers with 2-levels of anonymization
    - Los Nettos (an LA regional net)
    - FrontRange GigaPop (Colorado academic)
    - ServePath (San Jose commercial)
    - => eventually plan to allow user-provided analysis code on our boxes
  - curating some datasets (DoS, etc.)
- active data collection (since 2003 and ongoing)
  - **IP address census**: ping the world (*all* allocated v4 addresses, every quarter or so)
  - **IP address survey:** ping *some* of the world, often (1% of v4, every 11 minutes, for 1 week)
- support from DHS (infrastructure) and NSF (analysis)

The Internet

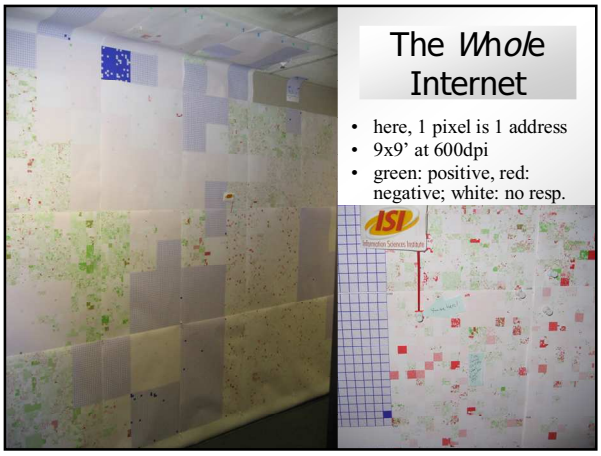
- average each /16
  - each pixel: 65k addresses
  - represents all 2^{32} addrs
- brightness: responsiveness
- green/red-ness: degree of positive vs. negative replies
- blue: areas not probed
- layout: Hilbert Curve





The *Whole* Internet

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



Caveats

- *not* a perfect statement of truth
  - misses NAT'ed hosts
  - misses non-ICMP-responsive hosts (those behind firewalls)
  - some pings are lost (we estimate <5%)
- *but* the *best current view* of the Internet;
  and a *new methodology* to be refined

"Your data is useless, everybody blocks pings" –common first reaction

"ghetto science" – slashdot "discussion"

We don't think so, and we have data to support our claim.

Are Pings Useful at All?

USC Survey (82k hosts)

category:		any	active
addresses probed	81,664		
non-responding	54,078		
responding any	27,586	100%	
ICMP or TCP	19,866	72%	100%
ICMP	17,054	62%	86%
TCP	14,794	54%	74%
Passive	25,706	93%	
ICMP only	656		
TCP only	1,081		
Passive only	7,720		

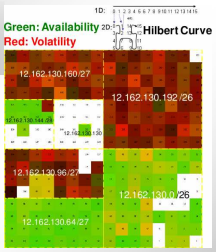
1M Random Addresses

category:		active
addresses probed	1,000,000	
non-responding	945,703	
responding either	54,297	100%
ICMP	40,033	74%
TCP	34,182	62%
both ICMP and TCP	19,918	
ICMP only	20,115	
TCP only	14,264	

YES! (given error estimates)
- USC says 24% low vs. solid baseline
- random Internet says ~25% low vs. weaker baseline
- ICMP strictly better than active TCP probing

What Good Are They?

- pretty pictures
- trying to get a handle on size of the Internet
- estimating *how* the net is used (work in progress with Xue Cai; figure at left)
- building a hit list of live edge hosts (ex: for Ark and other topology probers)



- estimating *how* the net is used (work in progress)
  - red is dial-up and dyanamic; green is servers

Additional LANDER Information

- http://www.isi.edu/ant/lander/
- part of PREDICT: http://www.predict.org
- all datasets are available

- active probing more info: see Heidemann et al.; "Census and Survey of the Visible Internet", ACM IMC 2008; doi:10.1145/1452520.1452542