Long-term Data Collection and Analysis of Outages at the Edge

John Heidemann
joint work with Lin Quan and Yuri Pradkin

8 February 2013
CAIDA Workshop on Active Internet Measurements
San Diego, California, USA

work supported by DHS S&T, Cyber Security Division

1. How Pings Measure Internet Outages?

PING 8.8.8.8 (8.8.8.8) 56(64)bytes of data.
64 bytes from 8.8.8.8: icmp_seq=1 ttl=251 time=99.6 ms
64 bytes from 8.8.8.8: icmp_seq=2 ttl=251 time=99.6 ms
64 bytes from 8.8.8.8: icmp_seq=3 ttl=251 time=99.6 ms
64 bytes from 8.8.8.8: icmp_seq=4 ttl=251 time=99.6 ms

-- 8.8.8.8 ping statistics --
3 packets transmitted, 3 received, 0% packet loss, time 200ms
rtt min/avg/max/mdev = 93.662/98.627/99.641/2.463 ms

2. What Role for Long-Term Data Collection?

- real data and new ideas
- collection for serendipity
- “instant” longitudinal study

=> what properties make successful long-term data collection?

Background: Active Probing with Pings

pings (ICMP echo request) draw positive replies when an IP address is in use

PING 8.8.8.8 (8.8.8.8) 56(64)bytes of data.
64 bytes from 8.8.8.8: icmp_seq=1 ttl=251 time=99.6 ms
64 bytes from 8.8.8.8: icmp_seq=2 ttl=251 time=99.6 ms
64 bytes from 8.8.8.8: icmp_seq=3 ttl=251 time=99.6 ms
64 bytes from 8.8.8.8: icmp_seq=4 ttl=251 time=99.6 ms

-- 8.8.8.8 ping statistics --
3 packets transmitted, 3 received, 0% packet loss, time 200ms
rtt min/avg/max/mdev = 93.662/98.627/99.641/2.463 ms

Background: Active Probing with Pings

pings (ICMP echo request) draw positive replies when an IP address is in use

PING 8.8.8.8 (8.8.8.8) 56(64)bytes of data.
64 bytes from 8.8.8.8: icmp_seq=1 ttl=251 time=99.6 ms
64 bytes from 8.8.8.8: icmp_seq=2 ttl=251 time=99.6 ms
64 bytes from 8.8.8.8: icmp_seq=3 ttl=251 time=99.6 ms
64 bytes from 8.8.8.8: icmp_seq=4 ttl=251 time=99.6 ms

-- 8.8.8.8 ping statistics --
3 packets transmitted, 3 received, 0% packet loss, time 200ms
rtt min/avg/max/mdev = 93.662/98.627/99.641/2.463 ms

Broader Goal for Outages: Understanding Edge Networks

- quickly know the impact of natural disasters
  - Hurricane Sandy, Tohoku Earthquake 2011, etc.
  - and human ones — like Egypt 2011, etc.

- learn about outage shapes
  - wide outages: many people
  - long outages: long time
  - and both

- in edge networks (24 address blocks, like 1.2.3.*)
  - most outages are small, inside ISPs, not from routing
  - e.g. J.Bush et al. (DMC 2007): ~50% smaller than notable prefixes
  - want to characterize what people see at home

Background: Active Probing with Pings
Pings Tell You Something But Not Everything

- Positive: block is up
- Negative: block is down or computer crashed, laptop suspended, computer address reassigned, probe or reply lost, firewall enabled
- Negative replies are ambiguous

So We Probe Multiple Addresses

- All negative together disambiguates: network is really down

Approach: Detect Changes in Ping Response

1. Probe multiple addresses in each block frequently
2. Gaps indicate block-level outages
3. Show block one per line; order to cluster by similarity

Approach: Detect Changes in Ping Response

1. Probe multiple addresses in each block frequently
2. Gaps indicate block-level outages
3. Show block one per line; order to cluster by similarity

Approach: Detect Changes in Ping Response

- Outages due to Hurricane Sandy
- Show block one per line; order to cluster by similarity
- Time 2014-10-01 to 2015-01-01
**Long-term Data: Internet Surveys**

- Internet Surveys
  - sample: 41k blocks (~2%) of active address space
  - half the same from survey to survey
  - half vary, with one-quarter chosen new each time
  - probe for 2 weeks
  - all addresses in each block every 11 minutes
- long-term effort
  - 31 sets to date (412 to 451), 21 from ≥2 sites, 9 from 3 sites
  - started in 2006, and ongoing through today...
- details and data are available
  - ISI-TR-678b: http://www.isi.edu/~johnh/PAPERS/Quan12a.html
  - data: http://www.isi.edu/ant/traces/

**Data About Sandy**

- look at one dataset: internet_address_replobing_1006-20121027
- 41,582 /24 blocks
- 11,900 geolocate to US
- 4,117 have enough response to analyze
- 60 of these don’t have states

**Outages at Sandy Landfall**

cluster of outages, starting with landfall

- marginal distribution to quantify impact

**Measuring the Impact**

after Sandy: U.S.-level of outages doubled to about 0.4% (compare daily median, blue line, before and after)

always some outages in US: about 0.2%

(proportional to amount of U.S. outage for 12 minutes)

back to baseline after about 4 days

**Where Are Outages? NY/NJ**

geolocation shows outage increase due from New York/New Jersey

**The Northeast, by Day**

3 days before Sandy landfall

4 days after Sandy landfall
**Role of Long-Term Data in Developing Outage Detection**

- real data and new ideas
- collection for serendipity
- “instant” longitudinal study

**Real Data is Inspirational**

- outage discovery idea came from raw data
  - gee, what’s that black vertical line?
  - how can we remove that error?
  - hmm... is this something deeper?

**Serendipity: Prominent and Unknown**

- Jan. 2011 Egyptian Revolution
- Jan. 2011 Australian Outage
- Mar. 2011 Japanese Earthquake
- Mexico (AS1982)
- AT&T and Comcast

- Our goal: small and big

**“Instant” Longitudinal Study**

- what’s “typical” in the Internet?
- do results vary by site?
- these are “free” if you already have the data

**Needs of Longitudinal Data**

- regular collection
- consistent, documented methodology
- careful archival
  - checksums (that you check!)
  - backups
- sharing
  - distribution procedures
  - results and what you learn (we have a wiki)

  => a non-trivial amount of work

**Cost of Long-term Data Collection**

- complaints
- traffic on target
  - survey puts 1 probe / 3 seconds
  - (1400 probes/hour) per block
- optimization can do **much** better
  - we’re scaling up outage detection to the whole analyzable Internet: 3.4M blocks
  - optimized probe rate <20 probes/hour per block
- but specialization incompatible with long-term, general-purpose datasets
What Next?

- outage detection from pings works
  - exciting what we see in old data
  - data enabled progress towards Internet-wide detection
- longitudinal data collection important
  - takes care and persistence

- blog: http://ant.isi.edu/blog/
- papers: http://www.isi.edu/ant/pubs
- datasets: http://www.isi.edu/ant/traces
  - can they enable your research idea?