Internet Outages: Reliability and Security

John Heidemann[1]

with Joseph Bannister[1,3*], Genevieve Bartlett[1], Christos Papadopoulos[2], Yuri Pradkin[1], Lin Quan[1,4*] , Abdulla Awabel[1] , Guillermo Baltra[1] , Dominik Staros[1]

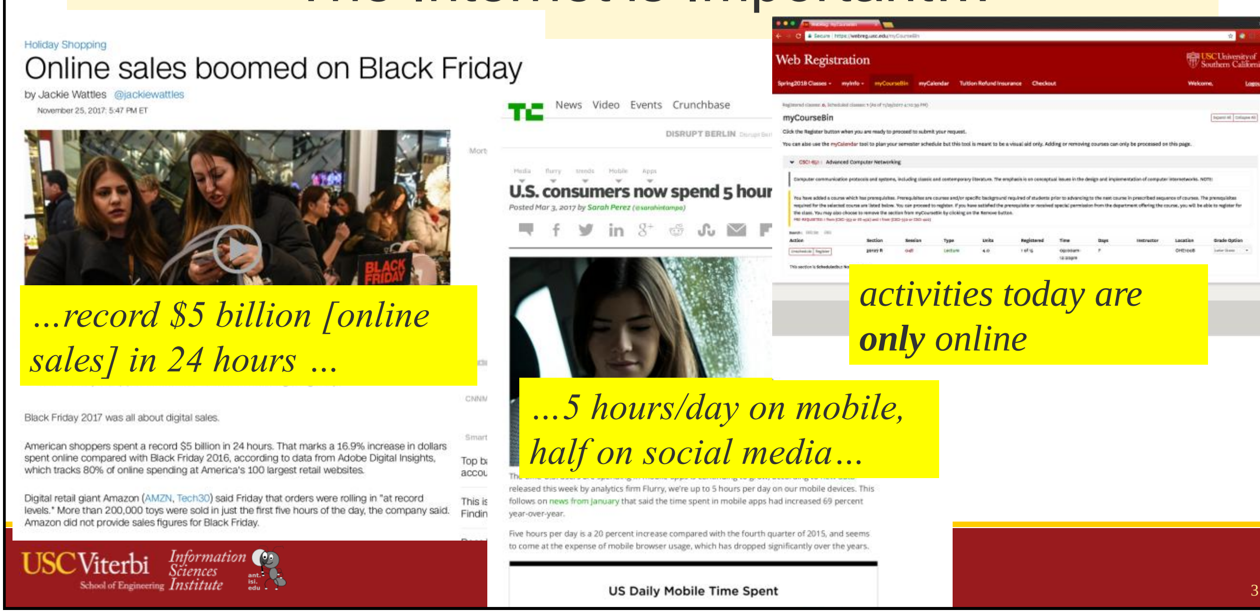
[1:] USC/ISI, [2:] Colorado State University, [3*:] Aerospace Corp, [4*:] Bank of China (* work done when at USC/ISI)

23 April 2018

This research is sponsored by the Department of Homeland Security (DHS) Science and Technology Directorate, HSARPA, Cyber Security Division, BAA 11-01-RIKA and Air Force Research Laboratory, Information Directorate under agreement number FA8750-12-2-0344, and contract number D08PC75599. The U.S. Gov't is authorized to reproduce and distribute reprints for Gov't purposes notwithstanding any copyright notation thereon. The views herein are those of the authors and do not necessarily represent those of DHS or the U.S. Gov't.

USC Viterbi School of Engineering — Information Sciences Institute

The Internet is Important…



…record $5 billion [online sales] in 24 hours …

*activities today are **only** online*

…5 hours/day on mobile, half on social media…

The *World* Is Important

hurricanes, floods, fires, blizzards…

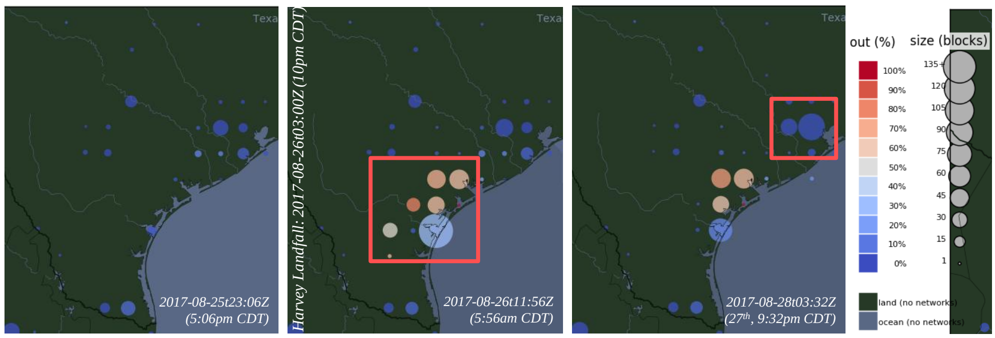
before landfall: **few outages**

serious outages (red circles), N. of Corpus Christi

many outages (large circles), in Houston-flooding

Hurricane Harvey, August 2017

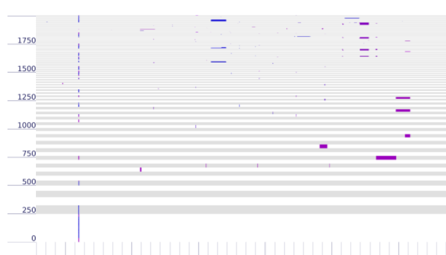
animation: (play) https://ant.isi.edu/outage/ani/harvey/



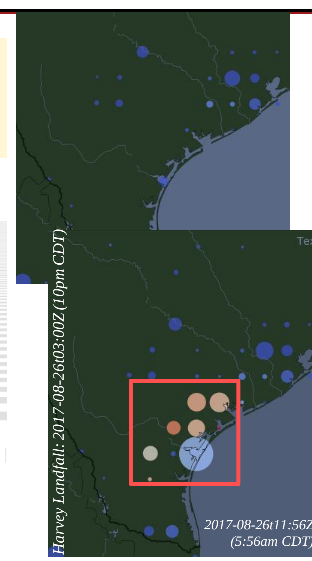
Network Reliability as Security

the Internet is important: Internet **reliability** is one aspect of **security**

communication without **intentional network interference**

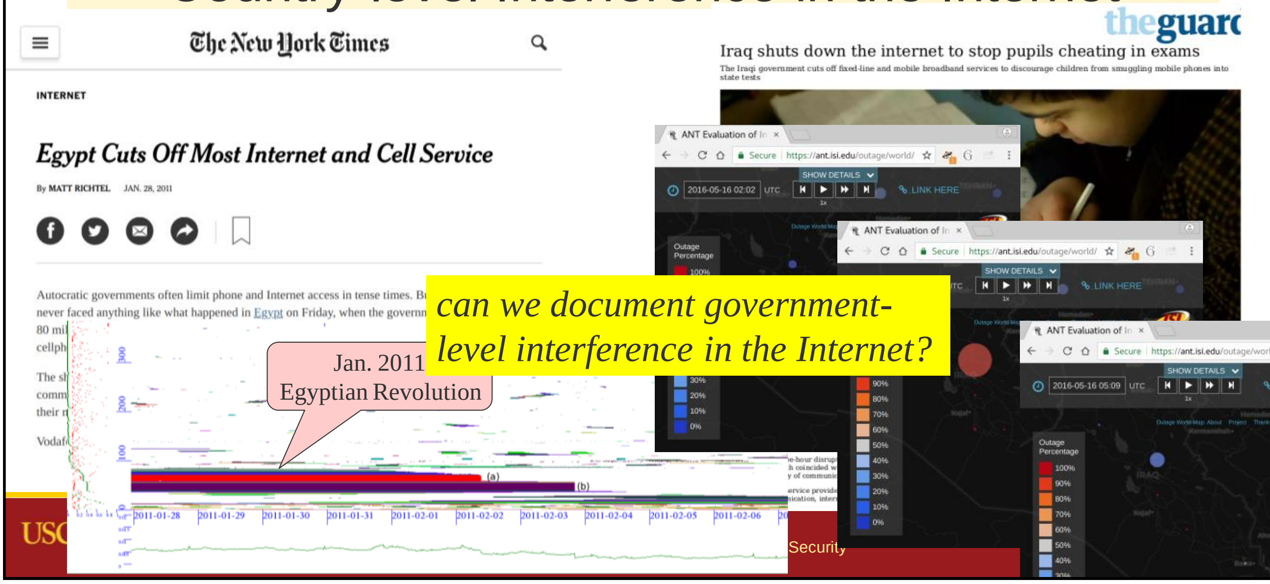


threats from **weaknesses in critical infrastructure**

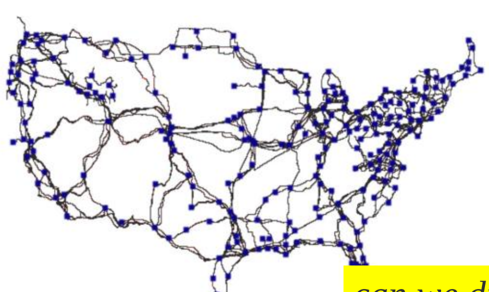


speedy **physical recovery to natural disasters**

Network Reliability as Security: Country-level Interference in the Internet

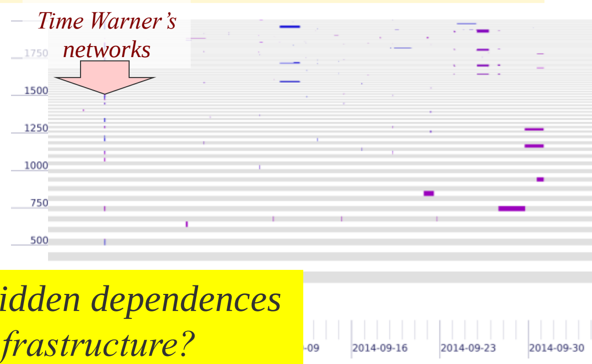


Network Reliability as Security: Infrastructure Resilience and Points-of-Failure



Physical conduits used by the U.S. I

From "InterTubes: A Study of the US Long-Haul Fiber-optic Infrastructure" by Durairajan, Barford, Sommers, and Willinger, ACM SIGCOMM, Aug. 2015



can we discover hidden dependences in the Internet's infrastructure?

Clustering algorithms discovering Time Warner's network from their Sept. 2014 outage.

Network Reliability as Security: Safety in the Physical World

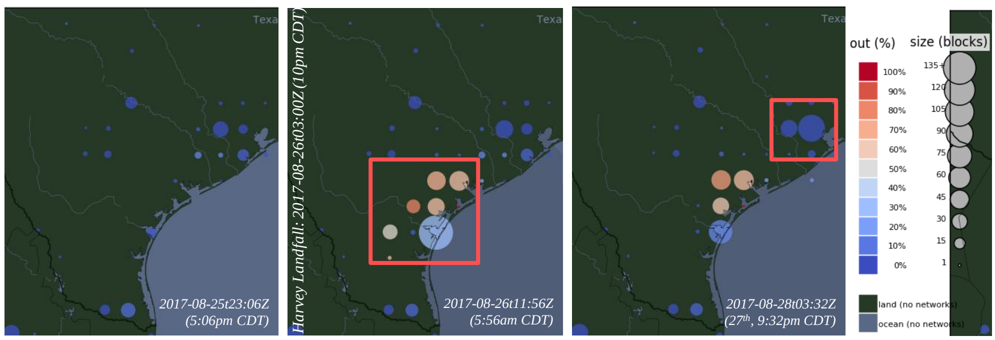
hurricanes, floods, fires, blizzards…

before landfall: **few outages**

serious outages (red circles), N. of Corpus Christi

many outages (large circles), in Houston-flooding

Hurricane Harvey, August 2017



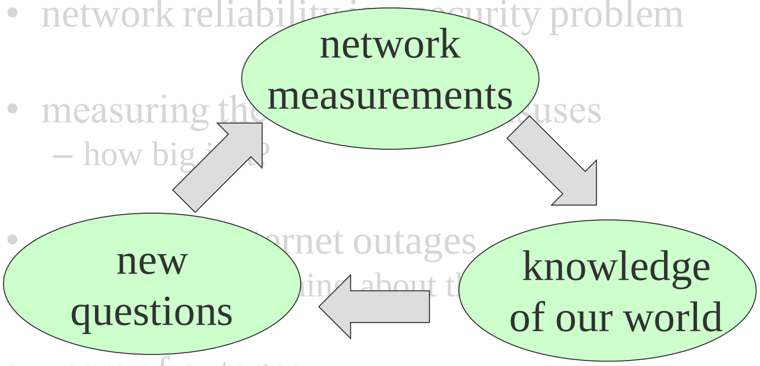
animation: (play)
https://ant.isi.edu/outage/ani/harvey/

Three Steps

- network reliability is a security problem
- measuring the Internet… Censuses
  – how big is it?
- measuring Internet outages
  – they say something about the real world, too!
- years of outages
  – revealing hidden dependencies

Three Steps

- network reliability is a security problem
- measuring the ... uses
  - how big is ?
- ...rnet outages
  - ...ing about t...
- years of outages
  - revealing hidden dependencies



results with scientific rigor

knowledge and data that others build on

Who We Are

outage detection and visualization

Guillermo Baltra, USC/ISI; Lin Quan, USC/ISI (now: Bank of China); **John Heidemann, USC/ISI, PI**; Yuri Pradkin, USC/ISI; Dominik Staros, USC/ISI and imaginewc.com

part of the LACANIC project: https://ant.isi.edu/lacanic/

Christos Papadopoulos, CSU co-PI; Hang Guo, USC/ISI; Abdul Qadeer, USC/ISI; Wei Lan, USC/ISI; Han Zhang, CSU; Liang Zhu, USC/ISI

USC Viterbi School of Engineering Information Sciences Institute

Information Sciences Institute

Colorado State University

with hosting from
USC/ISI—Marina del Rey and Arlington, VA
CSU
Keio University, Japan
Athens U. of Economics and Business
SurfNet, Netherlands

and ongoing collaboration with FCC to evaluate technology

LACREND is part of the DHS IMPACT program

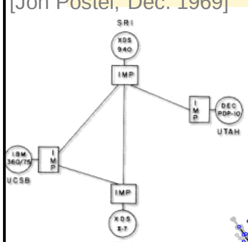
Impact Cyber Trust
www.impactcybertrust.org

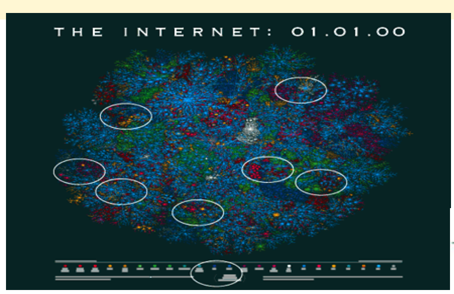
Three Steps

- network reliability is a security problem
- **measuring the Internet… Censuses**
  - **how big is it?**
- measuring Internet outages
  - they say something about the real world, too!
- years of outages
  - revealing hidden dependencies

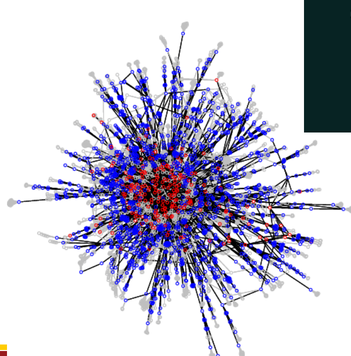
The Internet

[Jon Postel, Dec. 1969]



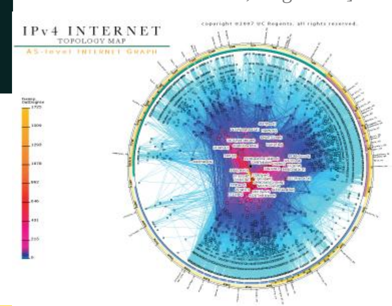


[map by CAIDA; data from Cheswick and Burch; 2000]



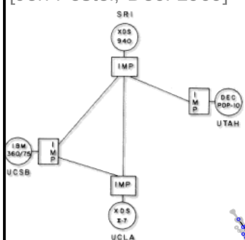
[Cable and Wireless (only); 1999, by Ramesh Govindan]

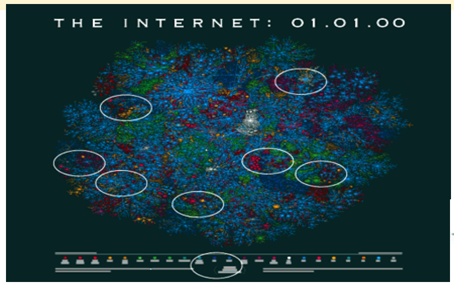
[AS-level map; CAIDA, Aug. 2007]



The Internet

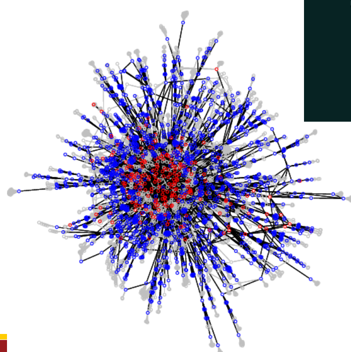
[Jon Postel, Dec. 1969]

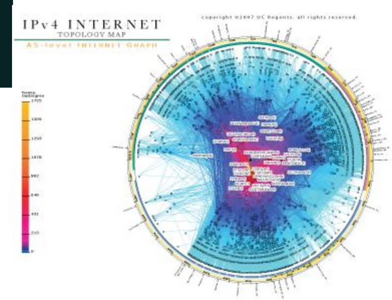




[map by CAIDA; data from Cheswick and Burch; 2000]

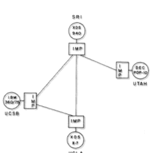
[AS-level map; CAIDA, Aug. 2007]

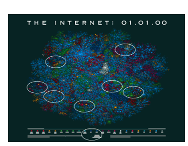


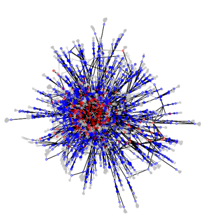


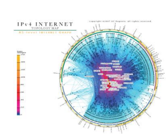
[Cable and Wireless (only); 1999, by Ramesh Govindan]k

The Internet

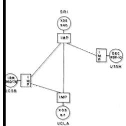




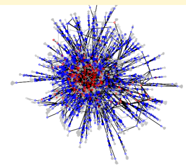


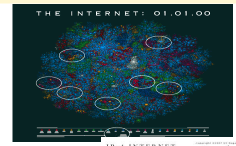


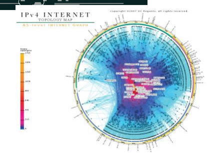
Prior Work: Studying Only The Core



map **all edge hosts**
(each public, unicast IPv4 addr)







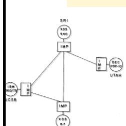
and scale to the **size of today's Internet**

just the network core

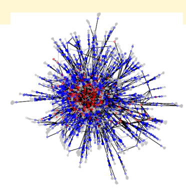
traceroute to each network; map routers and links
(Cheswick and Burch 2000; Tangmunarunkit et al, 2001; Spring et al, 2002)

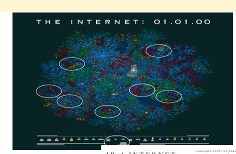
observe routing tables
(Huffaker et al, 2001; Meng et al, 2001; Francis et al, 2001)

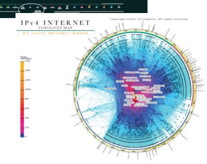
Us: Study *All* of *Today's* Internet



map **all edge hosts**
(each public, unicast IPv4 addr)







and scale to the **size of today's Internet**

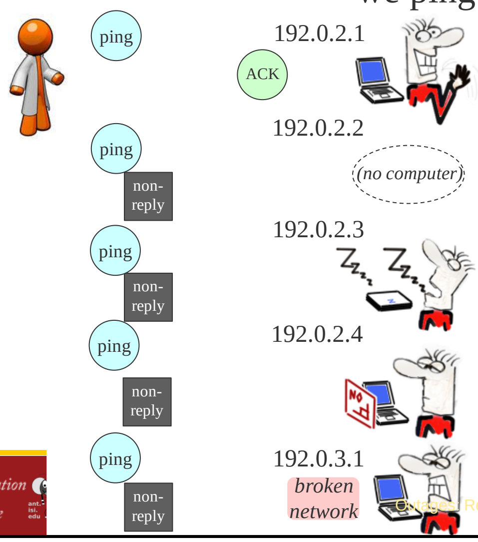
our approach:

ping *all* addresses once (census)
some addresses many times (survey)
quantify sources of error

goal: scientific impact
(ex: Q: how big is the Internet?)

Active Measurement of the Internet

we ping (ICMP echo request) all IPv4



we find computers

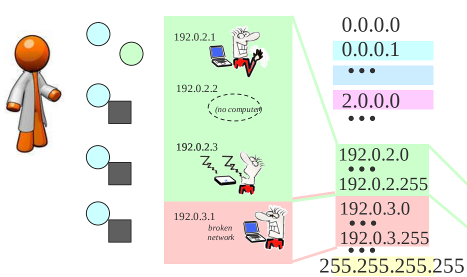
and unused space

temporarily unused: (sleeping)

intentionally silent (firewalled)

and *network* ***problems***

Our Insight: Observing the Internet Informs



we can probe the **whole Internet**

(the public, unicast, IPv4 Internet)

"only" 4 billion possible addresses
we probe **blocks**: adjacent 256 addrs
(like 192.0.2.0 to 192.0.2.255)
a census: all **12 million blocks**

challenge: **interpreting the results** and **probing sustainably**

scientifically rigorous results (known accuracy!)

minimal traffic for 24x7 coverage
no harm to the net (or annoying users!)

About Internet Addressing

IPv4 **addresses**
(today's Internet)
2^{32} addresses (~4 billion)
usually written: 4 parts, each 8-bits
192.0.2.1

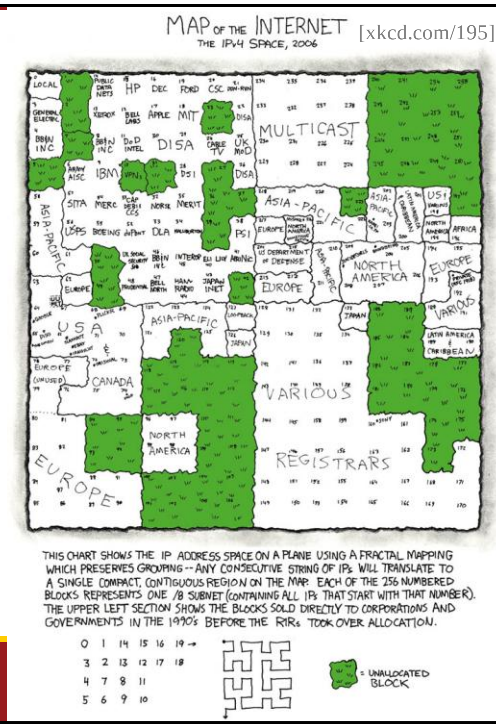
map linear addresses

1D: 0 1 2 3 4 5 6 7…

etc.

to 2D Hilbert Curve

2D:
0 1 14 15
3 2 13 12
4 7 8 11
5 6 9 10



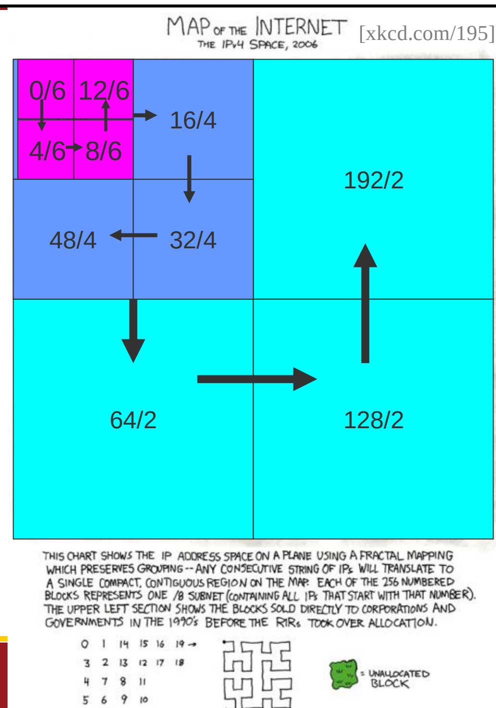
[xkcd.com/195]

About Internet Addressing

IPv4 **addresses**
(today's Internet)
2^{32} addresses (~4 billion)
usually written: 4 parts, each 8-bits
192.0.2.1

address **blocks:** adjacent
addresses with same first *n* bits
192.0.*.* /16
or just 192.0/16
(prefix=192.0, n=16)

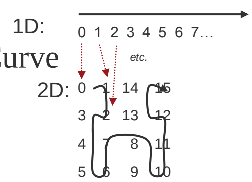
blocks are squares on map

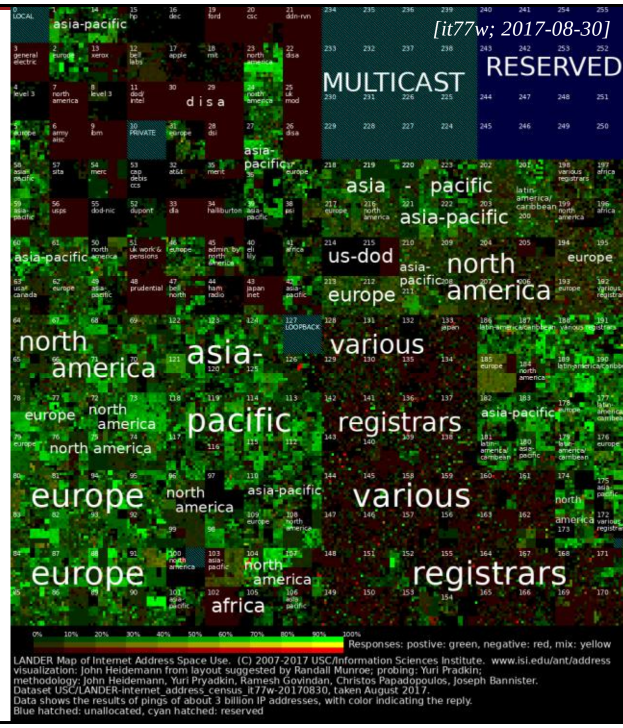


[xkcd.com/195]

The Internet

- each pixel is 65k IP addresses (a /16)
- 65k pixels = all 2^{32} addresses
- brightness: responsiveness
- green/red-ness: degree of positive vs. negative replies
- blue: reserved, not probed
- cyan: private or multicast, not probed
- layout: Hilbert Curve





23

[data: it44w taken Nov. 2011]

The *Whole* Internet

- here, 1 pixel is 1 address
- 2.8x2.8m (9x9') at 600dpi
- green: positive, red: negative; white: no resp.

24

But Does This Mean *Anything?*
(validation!)

- *not* a perfect statement of truth
  - misses NAT'ed hosts
    (Network Address Translation)
  - misses non-ping-responsive hosts (from firewalls)
  - some pings are lost (we estimate <5%)
- the *best current view* of the Internet; and a *new methodology* to refine data suggesting *new questions*

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

"ghetto science" –slashdot "discussion"

We disagree, and our data supports our claim.

Sources of Error

- overcounting
  - routers and multi-homed hosts: *estimated at <6% in paper*
- undercounting
  - probe loss: *random due to probe order; use 1-repair process to recover single losses in survey*
  - firewalled hosts: *coming up*
- variance
  - measurement location: *doesn't matter; normal error*
  - sampling error:
    - *can predict from theory*
    - *function of probe frequency*
    - *surveys within 0.4% (with 95% confidence)*
  - births/deaths during survey: *estimate in paper*
  - probe type (ICMP vs. TCP): *ICMP consistently more complete*

Sources of Error

- overcounting
  - routers and multi-homed hosts: *estimated at <6% in paper*
- undercounting
- variance
  - measurement location: *doesn't matter; normal error*
  - sampling error:

Validation:

method: compare ICMP (pings), TCP, and observed traffic

- with **USC's network** (good ground truth, but maybe biased)
- with **million random IP addresses** (weaker ground truth but unbiased)

details: "Census and Survey of the Visible Internet", Heidemann, et al.; ACM IMC, Oct. 2008

Validating with USC and a Random Sample

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		

Census is *incomplete*, but can *estimate* error

=> we see 62% of truth at USC

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	14,264	

Both USC and random sample are similar

⇒ 62% or 74% of truth

⇒ USC seems representative

responding any: addresses at our border routers, or in ICMP or TCP scans

Impact of IPv4 Censuses

- how big is the net?
- complete allocation of IPv4: but are we using it?
- how do we use the net?
- assisting topology discovery?
- what about network reliability?
- do others build on it?

Impact of IPv4 Censuses

- how big is the net?
- complete allocation of IPv4: but are we using it?

"Census and Survey of the Visible Internet", Heidemann, Pradkin, Govindan, Papaopoulos, Bartlett, Bannister; ACM IMC, Oct. 2008

- how do we use the net?

"Understanding Block-level Address Usage in the Visible Internet", Cai & Heidemann; ACM SIGCOMM, 2010

- assisting topology discovery?

"Selecting Representative IP Addresses for Internet Topology Studies", Fan & Heidemann; ACM IMC, Nov. 2010

- what about network reliability?

"Trinocular: Understanding Internet Reliability Through Adaptive Probing", Quan & Heidemann, SIGCOMM,.2013

- do others build on it?

datasets use and follow-on work by others

Impact of IPv4 Censuses

- how big is the net?
- complete a
  but are we
- how do we use the net?
- assisting t
- what abou
- do others l

“Census and Survey of the Visible Internet”, Heidemann, Prad
Govindan, Papadopoulos, Bartlett, Bannister; ACM IMC, Oct. 2

“Understanding Block-level Address Usage in the Visible Inter
& Heidemann; ACM SIGCOMM, 2010

ve IP Addresses for Internet Topology S
M IMC, Nov. 2010

ding Internet Reliability Through Adapt
emann, SIGCOMM, 2013

-on work by others

foundations

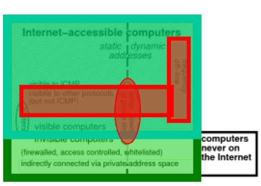
current mapping better

new: net reliability

use by others

use by others

How Big is the Internet?

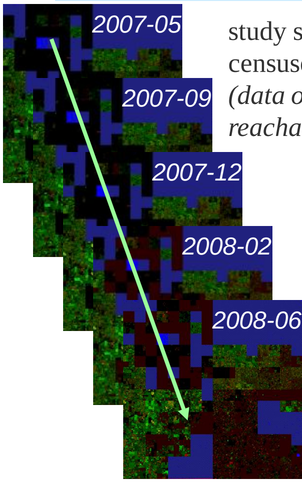


(correction procedure for estimated error)

data: Aug. 2017
(census *USC/LANDER internet_address_survey_it77w-20170830)*

address type	count	%IPv4	%uni-cast
IPv4 addresses	4,293M	100%	
special (multicast, pvt, etc.)	587M	13%	
unallocated	5.9M	10%	
allocated unicast	3,702M	86%	100%
responsive	419M	9%	11%
positive	371M	8%	10%
negative	50M	1%	1%
non-responsive	3,281M	76%	89%
best estimate: in-use, allocated unicast (scale by 1.6-1.9)	**670M-796M**	**17%**	**19%**

IP Hitlists
(Where Should Topology Studies Probe?)

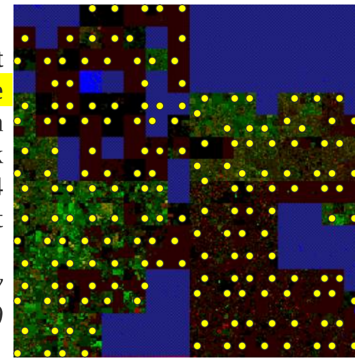


study series of censuses *(data on all reachablity)*

look at each /24's history

find **best representative** for each /24 block over the IPv4 Internet

(best: likely to respond)



hitlists are available for research use; see https://ant.isi.edu/traces/

"Selecting Representative IP addresses for Internet Topology Studies", Fan & Heidemann, ACM IMC 2010

Others Build On It: Using Our Data and Redoing Our Code

- we run as service
- sharing with
  - https://www.impactcybertrust/
  - in US, Japan, Australia, UK, Canada, Netherlands, Israel, Singapore
  - directly if IMPACT cannot
- **data shared** as of 2018-03-31:
  - **1231 datasets** (1.3TB compressed)
  - **96 unique researchers**

several **groups scan IPv4, building on our work**

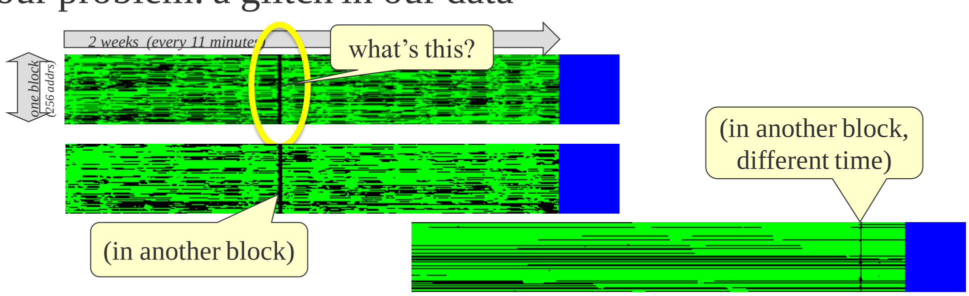
- Carna Botnet (2013-03)
  - anonymous grayhat hacker
  - reused 30k compromised home routers *and* some of our code
  - also scanned for services
  - (vs. us: *ethical* collection from known servers, only scans for presence)
- ZMap (2013-08)
  - Dumeric, Wustrow, Halderman (U.Mich), Usenix Security
  - inspired by our work
  - goal: *as fast as possible*
  - (vs. us: politely, at moderate rate)
- and MassScan (2014)

Three Steps

- network reliability is a security problem
- measuring the Internet… Censuses
  - how big is it?
- **measuring Internet outages**
  - **they say something about the real world, too!**
- years of outages
  - revealing hidden dependencies

Network Reliability

- our problem: a glitch in our data



- can we find and fix these (to get on with our real work)

… leads to the next part of this talk

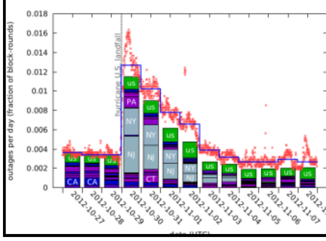
From Censuses to Outages

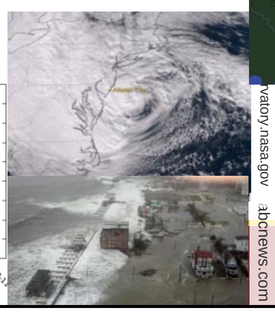
with interpretation
pinging the Internet
tells network outages
that tell about real world events

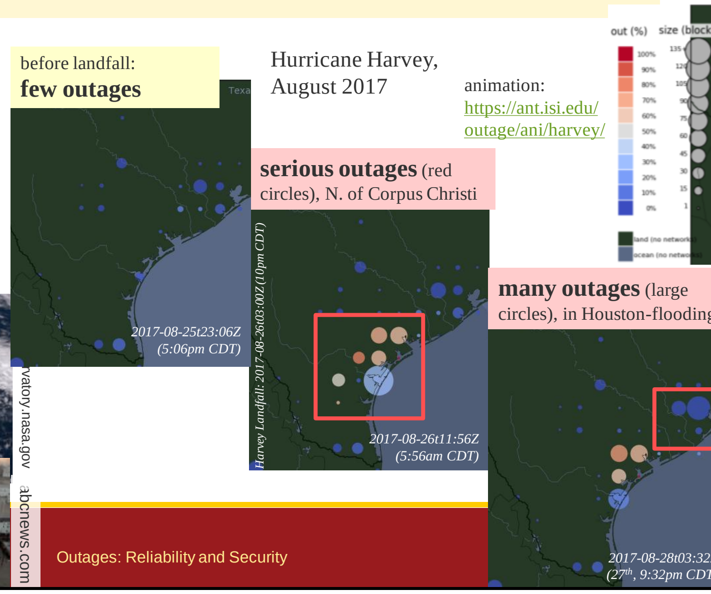
responding and stopping
…says something

time

net up *down*







Hurricane Harvey, August 2017

animation: https://ant.isi.edu/outage/ani/harvey/

before landfall: **few outages**

2017-08-25t23:06Z (5:06pm CDT)

serious outages (red circles), N. of Corpus Christi

2017-08-26t11:56Z (5:56am CDT)

many outages (large circles), in Houston-flooding

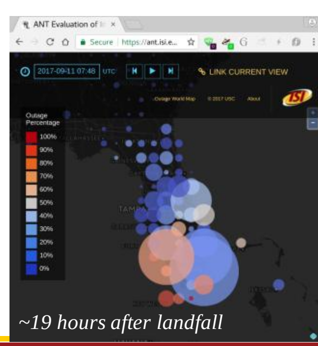
2017-08-28t03:32Z (27th, 9:32pm CDT)

Hurricane Irma: Watching Recovery

before, during and after disasters: Irma, Sept. 2017 in Florida…
good recovery underway 24 hours after landfall

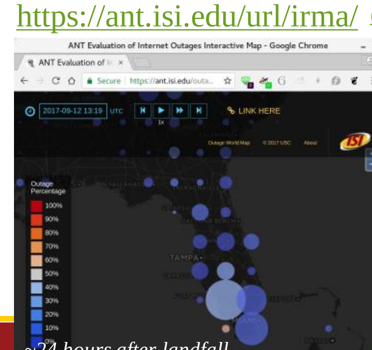
Irma landfall: 2017-09-10t13:10Z at Cudjoe Key, Florida

~12 hours after landfall



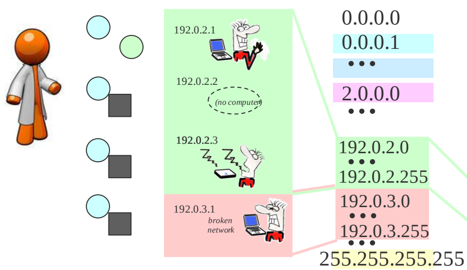
~19 hours after landfall

https://ant.isi.edu/url/irma/ (play)



~24 hours after landfall

Our Insight: Observing the Internet Informs



we can probe the **whole Internet**

(the public, unicast, IPv4 Internet)

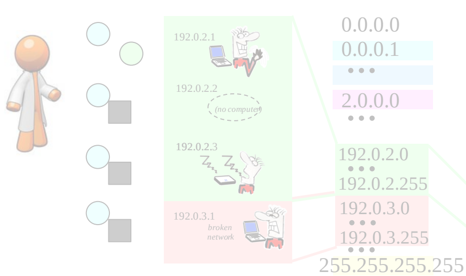
"only" 4 billion possible addresses
we probe **blocks**: adjacent 256 addrs
(like 192.0.2.0 to 192.0.2.255)
a census: all **12 million blocks**

challenge: **interpreting the results** and **probing sustainably**

scientifically meaningful results (known accuracy!)

minimal traffic for 24x7 coverage
no harm to the net (or annoying users!)

Our Insight: Observing the Internet Informs



we can probe the **whole Internet**

(the public, unicast, IPv4 Internet)

active probing
=> *known precision*

"only" 4 billion possible addresses
we probe **blocks**: adjacent 256 addrs
(like 192.0.2.0 to 192.0.2.255)
a census: all **12 million blocks**

challenge: **interpreting the results** and **probing sustainably**

scientifically meaningful results (known accuracy!)

observe blocks
=> *disambiguate*

minimal traffic for 24x7 coverage
no harm to the net (or annoying users!)

send just enough
=>*minimize cost*

Active Measurement of the Internet

we ping (ICMP echo request) all IPv4

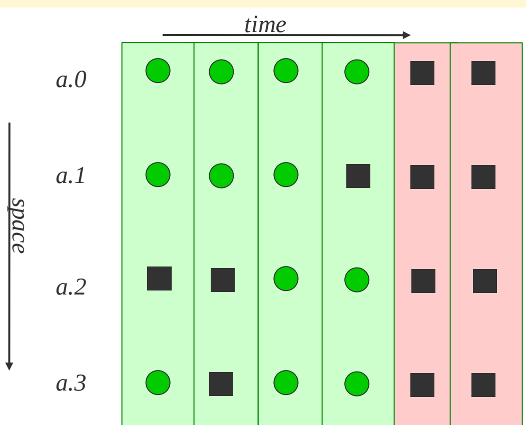
we find computers

challenge for outages: ambiguity in non-replies

intentionally silent (firewalled)

and *network* ***problems***

Observing Blocks to Disambiguate Replies



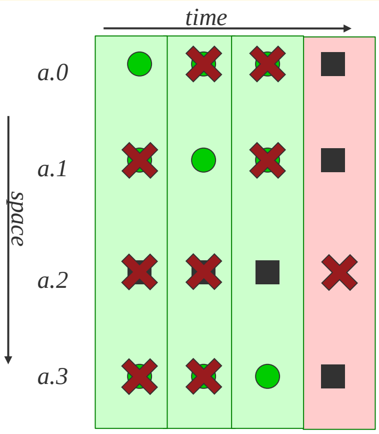
(blocks: really have 256 addresses, we show 4 here)

single negative: address is down
or
computer crashed
laptop suspended
computer address reassigned
probe or reply lost
firewall enabled

multiple probes
address ambiguity

all *negative: block is down* !!!

Probing Politely: *Just Enough*



polite: minimal traffic to your net

positive responses => block is up
but don't need all 4 to learn

1. instead: probe one by one
2. find **one is up** => **stop early**
3. if **try is down** => **try again**
   => **stop less early**
4. **several fail** => **block down**

adaptive probing uses Bayesian inference
informed by model of block response

probing politely =>
observing without harm

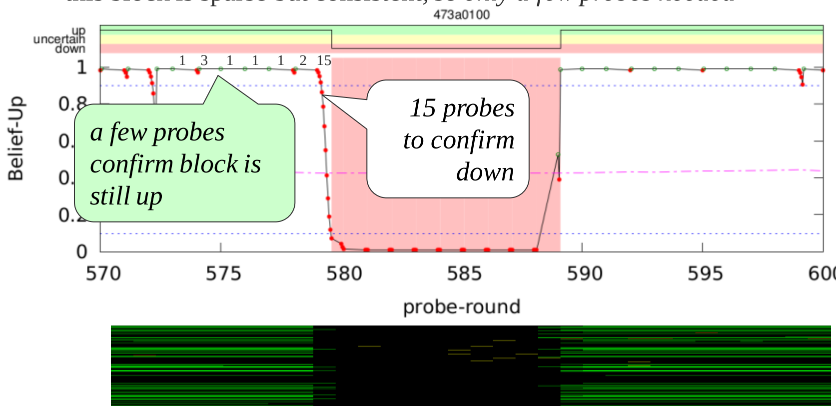
Key Properties of Trinocular

- Trinocular: active probing to detect Internet edge outages
  - **principled**: probe only when needed (informed by Bayesian inference)
  - **precise**: outage duration ±330s (half of probing interval)
  - **parsimonious**: only +0.7% background radiation (at target /24, per Trinocular instance)

(details: "Trinocular: Understanding Internet Reliability Through Adaptive Probing", Quan, Heidemann, Pradkin, SIGCOMM Aug. 2013)

Principled: Bayesian Inference Interprets Probes

model: every responding |E(b)|=111, active A(E(b))=0.515
this block is sparse but consistent, so *only a few probes needed*



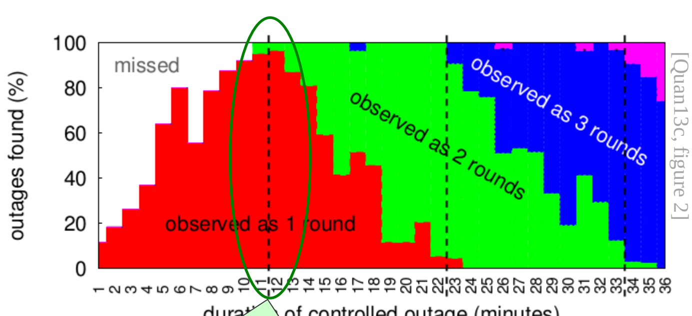
ground truth
(data for complete /24)

Modeling + Bayesian Inference says how many probes

probe result	prior U^*	$P(probe \mid U^*)$	reason
n	U	$1 - A(E(b))$	inactive addr.
p	U	$A(E(b))$	active addr.
n	$\bar{U}$	$1 - (1 - \ell)/\lvert b \rvert$	non-response to block
p	$\bar{U}$	$(1 - \ell)/\lvert b \rvert$	lone router?

$$B'(\bar{U}) = \frac{P(p \mid \bar{U}) B(\bar{U})}{P(p \mid \bar{U}) B(\bar{U}) + P(p \mid U) B(U)}$$

Precise: Detect All Outages?

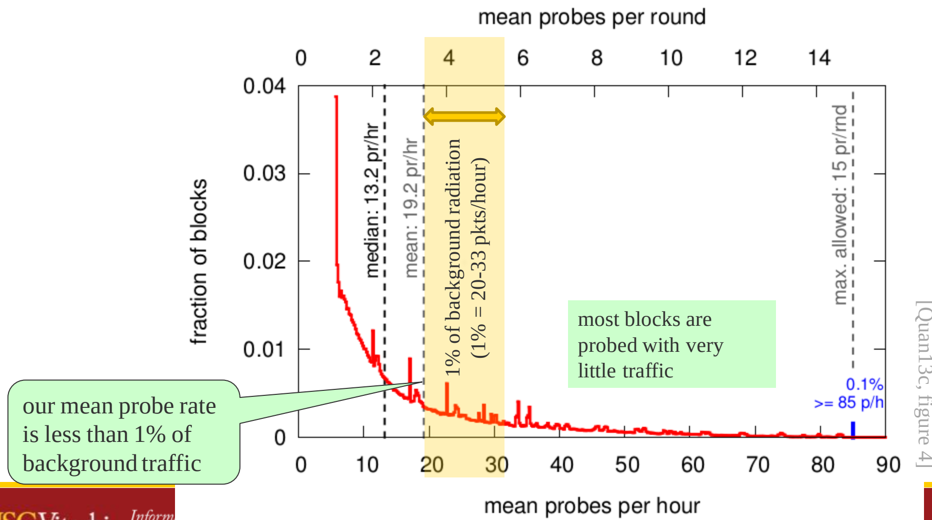


We detect **all** outages longer than 11 minutes (the probing interval)

Experiment:

Controlled outages (random duration, 1 to 36 minutes) in test block, measured from 3 different sites (2 in US, 1 in Japan).

Parsimonious: Probing Rate



Expriment:

Trinocular: post-facto analysis of 48 hours operation; background ration: from [Wustrow et al, ACM IMC 2010] ; today it is much higher

Impact of Outage Detection

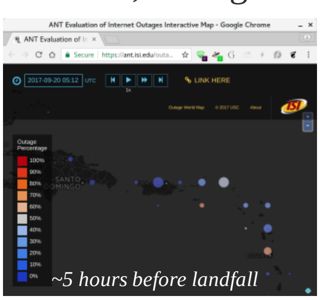
- quantified impact of hurricanes
- observed world events and policy
- others interested in data
- ongoing collaboration with FCC

Impact of Outage Detection

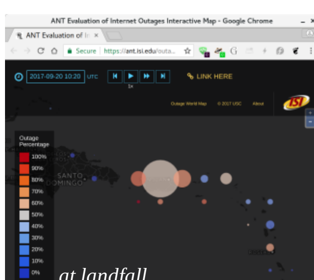
- **quantified impact of hurricanes**
  - **previously: Harvey (2017) and Irma (2017)**
  - **next: Maria (2017)**
- operational network outages
- relationship to government policy

Hurricane Maria: Watching Recovery

before, during and after disasters: Maria in Puerto Rico



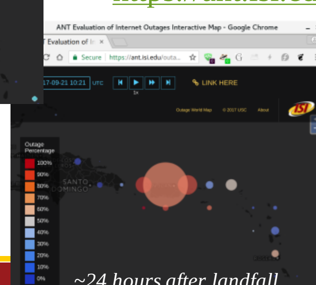
~5 hours before landfall



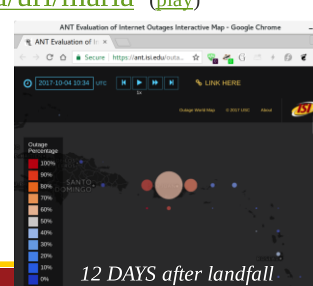
at landfall

Maria landfall:
2017-09-20t10:15Z
at Yaboucoa, P.R.

https://ant.isi.edu/url/maria (play)



~24 hours after landfall



12 DAYS after landfall

~19 hours after landfall

Impact of Outage Detection

- quantified impact of hurricanes
  - previously: Harvey (2017) and Irma (2017)
  - next: Maria (2017)
- **operational network outages**
- relationship to government policy

U.S. Outages: August 2014

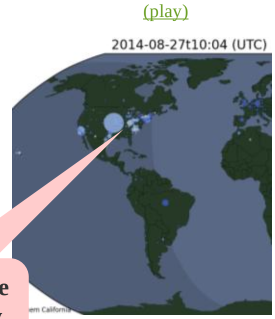
animating outages for the whole Internet

- this dataset:
  - 4M blocks
  - all of 2014q3
- events:
  - Time Warner outage on 2014-08-27 starting 9:20Z
  - ~11 million customers

--before

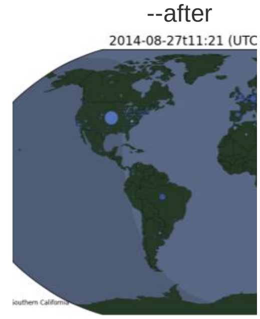
https://ant.isi.edu/outages/ani/tw/

(play)



--**during**

--after

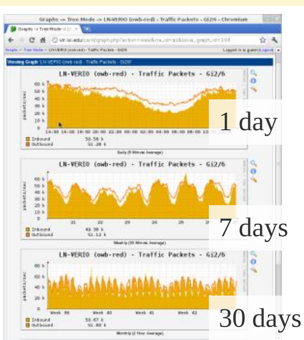


two hour outage affected nearly ~11M customers

Impact of Outage Detection

- quantified impact of hurricanes
  - previously: Harvey (2017) and Irma (2017)
  - next: Maria (2017) and Sandy (2012)
- operational network outages
- **relationship to government policy**

Does The Internet Sleep?



well known: traffic is diurnal
(seen *locally* everywhere)

people use computers

but not always

computers sleep, too

what about IPv4 **address usage**?
can we see **global view** ?

is there "more Internet" in the day?

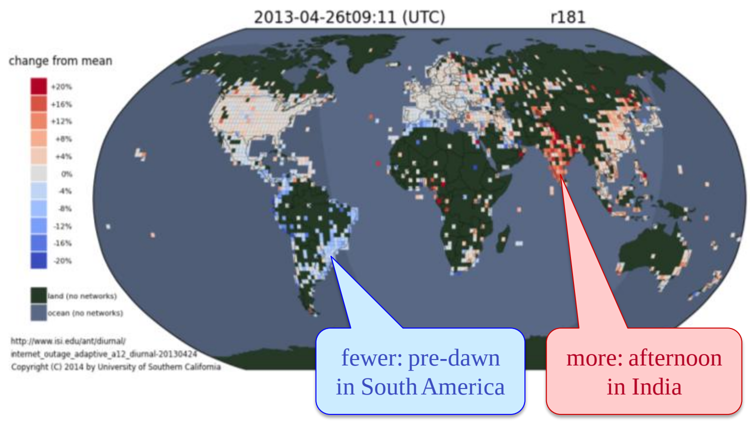
The Internet, Awake and Asleep

pinging the Internet shows *active addresses*

red: more than typical
white: typical
blue: fewer

parts of the Internet sleep:
more active during the day

https://ant.isi.edu/diurnal/ani/
(play)



Why Study "Sleep"?

sleep reflects policy

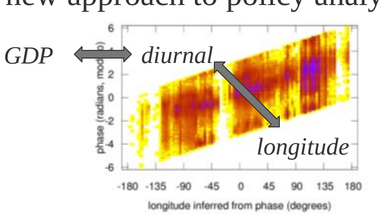
CONNECTING AMERICA: THE NATIONAL BROADBAND PLAN
U.S. FCC, March 2010

always-on networks a requirement for "broadband"

diurnal measures network maturity

sleep correlates with things

new approach to policy analysis



sleep affects outage detection

must not confuse sleep *with* down

how big is the net?

long-term goal

diurnal affects estimate

Three Steps

- network reliability is a security problem
- measuring the Internet… Censuses
  - how big is it?
- measuring Internet outages
  - they say something about the real world, too!
- **years of outages**
  - **revealing hidden dependencies**

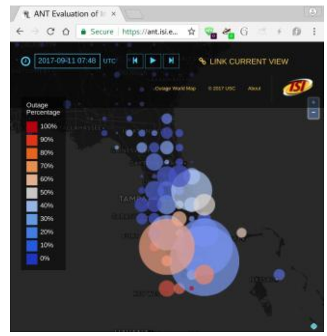
Analyzing Long-Term Data

- outage data, 24x7, since Nov. 2013
- about 40TB (!)
- about 20k observations x 4M blocks: 80G datapoints (!!)
- how to make sense of it?
  - interactive visualization
  - automated clustering

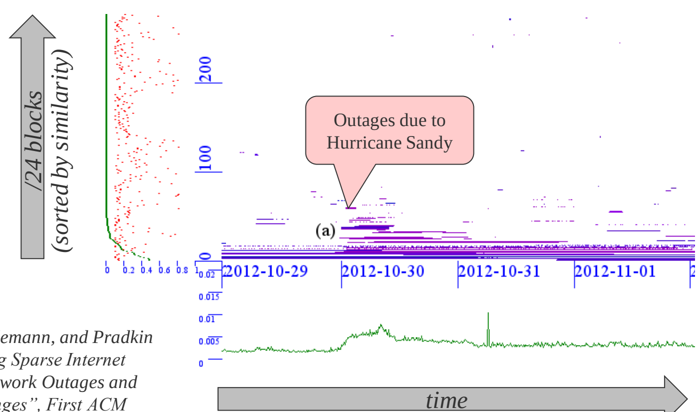
Geographic Visualization

- on the web: https://ant.isi.edu/outage/world/
- key features
  - circle size: *number* of blocks out
  - color: *percent* of blocks out
  - time selection
  - geographic zoom and pan
  - **geography: easy to relate to (what operators ask for!)**

Florida, ~19 hours after landfall of Hurricane Irma



Non-Geographic Visualizations: the *Network* in Outages

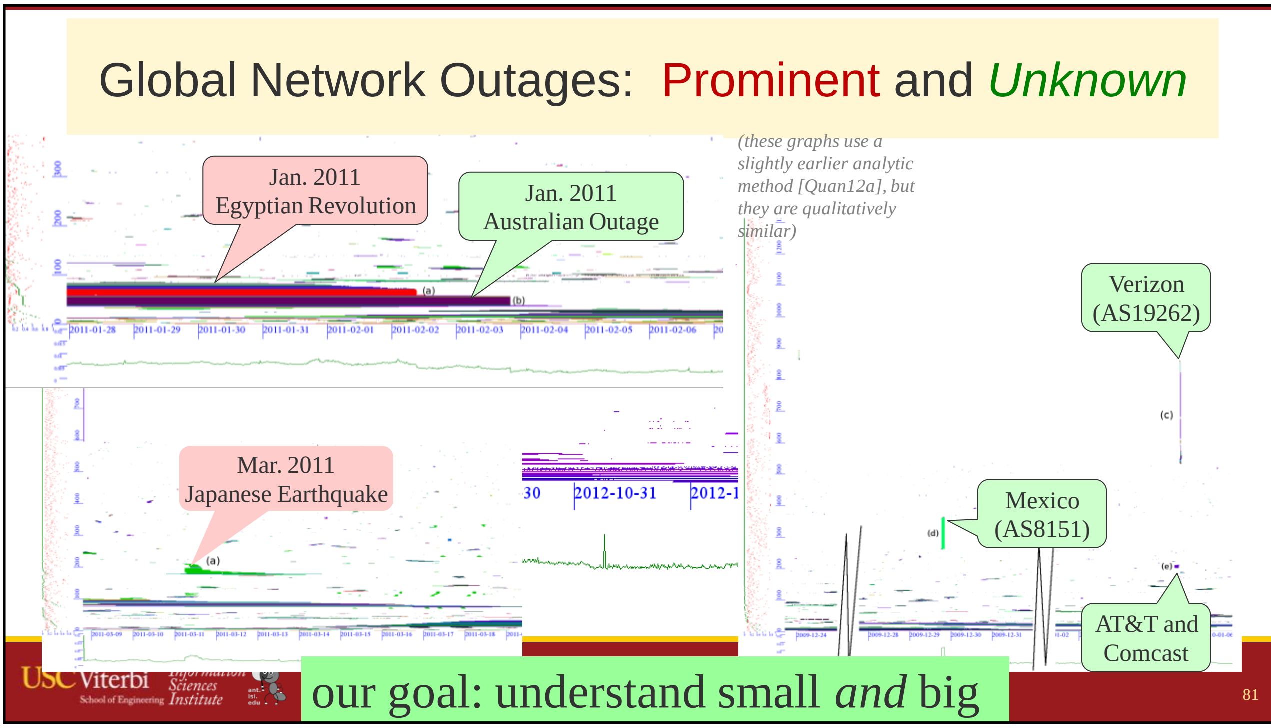


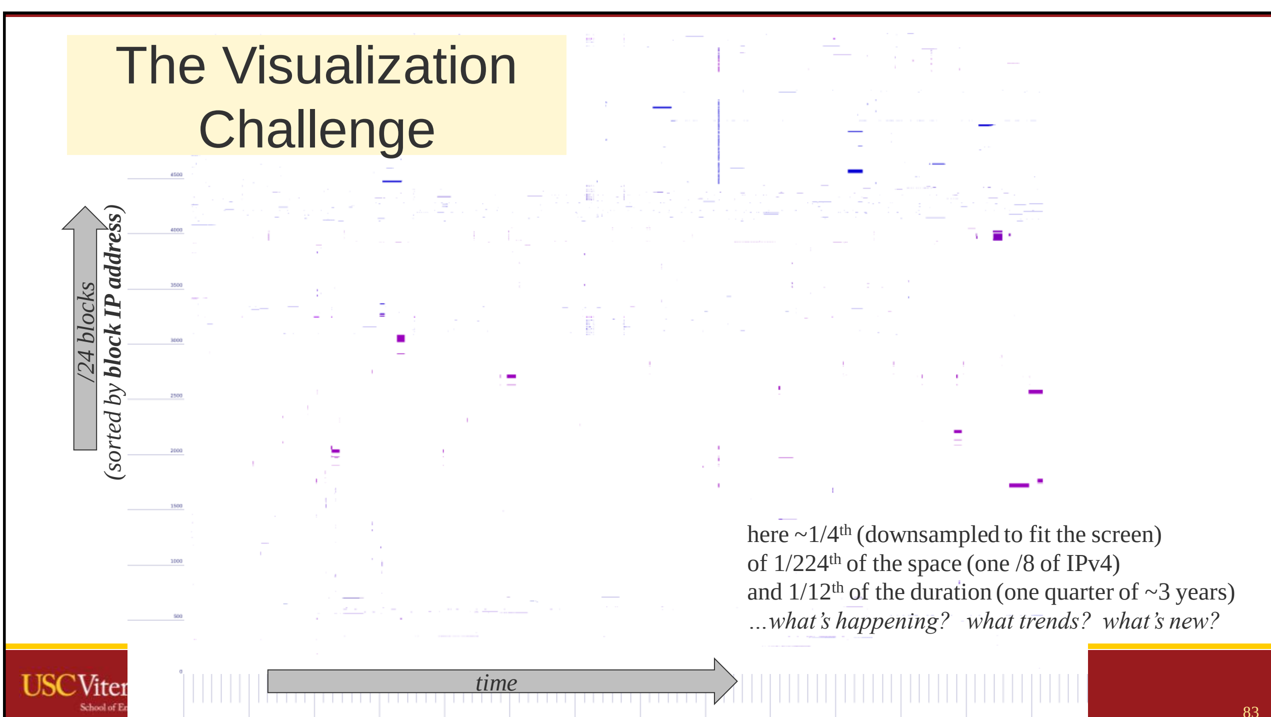
goal: reveal patterns

find dependencies among networks

(colored areas are outages, color shows location)

Quan, Heidemann, and Pradkin "Visualizing Sparse Internet Events: Network Outages and Route Changes", First ACM Workshop on Internet Visualization, Nov. 2012


Global Network Outages: Prominent and Unknown
(these graphs use a slightly earlier analytic method [Quan12a], but they are qualitatively similar)
Jan. 2011
Egyptian Revolution
Jan. 2011
Australian Outage
Mar. 2011
Japanese Earthquake
Verizon
(AS19262)
Mexico
(AS8151)
AT&T and
Comcast
our goal: understand small and big
81


The Visualization Challenge
/24 blocks
(sorted by block IP address)
time
here ~1/4th (downsampled to fit the screen)
of 1/224th of the space (one /8 of IPv4)
and 1/12th of the duration (one quarter of ~3 years)
...what's happening? what trends? what's new?
83

Efficient Visualization and Clustering

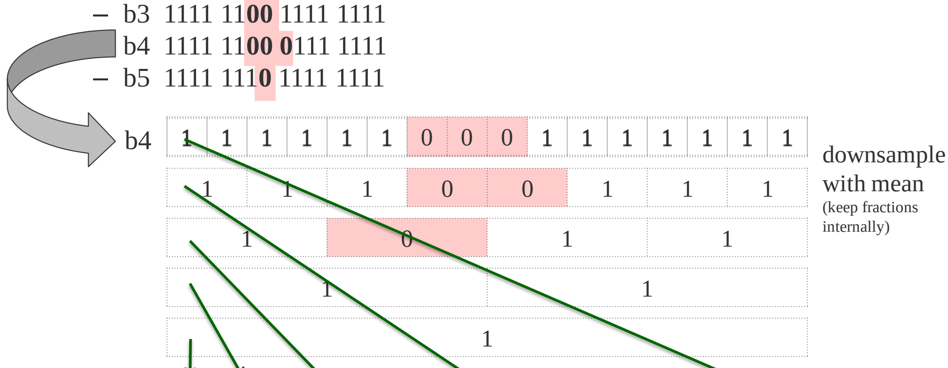
- **visualization with linear ordering algorithm**
  - runtime: O(n log n log m)
  - for n blocks and m duration timesteps
- approach:
  - map clustering to sorting: O(n log n) in time
  - sort on *multi-timescale bitmap:* O(log m) in space
- **event clustering**
  - runtime O(n^2)
  - parallelizes with Map/Reduce
- approach
  - find blocks that transition at the same time

Details in "Back Out: End-to-end Inference of Common Points-of-Failure in the Internet (extended)". ISI-TR-724, Feb., 2018. *www.isi.edu/~johnh/PAPERS/Heidemann18b.pdf*

Multi-Timescale for Similarity

- input: outage timeseries from 5 /24 blocks
  - b1 1111 111**0** 1111 1111
  - b2 1111 1111 1111 111**0**
  - b3 1111 11**00** 1111 1111
  - b4 1111 11**00** **0**111 1111
  - b5 1111 111**0** 1111 1111

goal: cluster by "similarity"



concatenate: 1 - 11 - 10**1**1 - 111**0** **0**111 - 1111 11**00** **0**111 1111

Multi-Timescale Mapping Results

- input: outage timeseries from 5 /24 blocks
  - b1 1111 111**0** 1111 1111
  - b2 1111 1111 1111 111**0**
  - b3 1111 11**00** 1111 1111
  - b4 1111 11**00** **0**111 1111
  - b5 1111 111**0** 1111 1111

goal: cluster by "similarity"

- apply to all blocks…
  - b1 1 - 11 - 1111 - 111**0** 1111 - 1111 111**0** 1111 1111
  - b2 1 - 11 - 1111 - 1111 111**0** - 1111 1111 1111 111**0**
  - b3 1 - 11 - 1**0**11 - 111**0** 1111 - 1111 11**00** 1111 1111
  - b4 1 - 11 - 1**0**11 - 111**0** **0**111 - 1111 11**00** **0**111 1111
  - b5 1 - 11 - 1111 - 111**0** 1111 - 1111 111**0** 1111 1111

Multi-Timescale Mapping Results

- input: outage timeseries from 5 /24 blocks
  - b1 1111 111**0** 1111 1111
  - b2 1111 1111 1111 111**0**
  - b3 1111 11**00** 1111 1111
  - b4 1111 11**00** **0**111 1111
  - b5 1111 111**0** 1111 1111

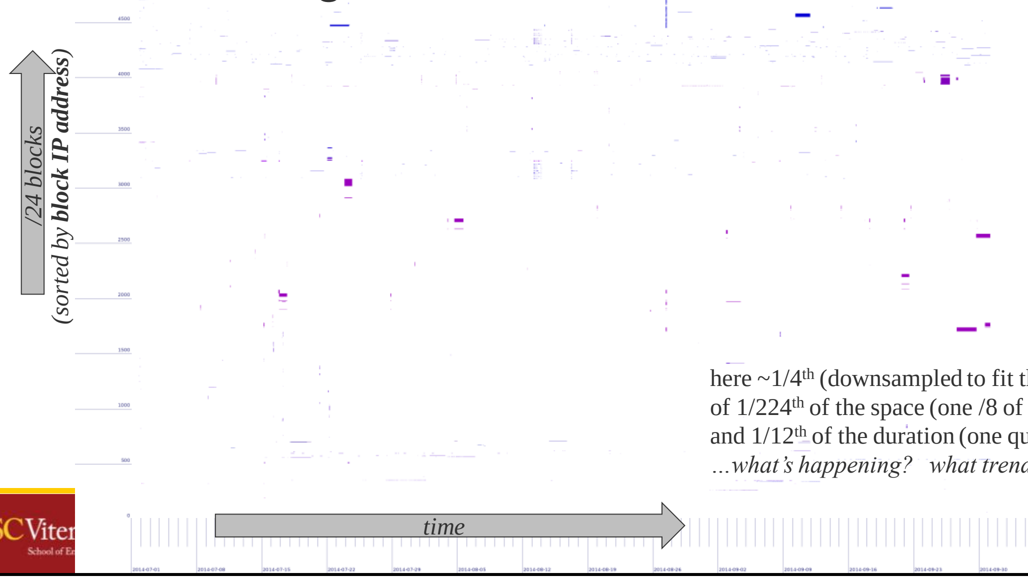
goal: cluster by "similarity"

define similar as adjacent in multi-timescale vectors

- apply to all blocks and **sort**
  - b2 1 - 11 - 1111 - 1111 111**0** - 1111 1111 1111 111**0**
  - b1 1 - 11 - 1111 - 111**0** 1111 - 1111 111**0** 1111 1111
  - b5 1 - 11 - 1111 - 111**0** 1111 - 1111 111**0** 1111 1111
  - b3 1 - 11 - 1**0**11 - 111**0** 1111 - 1111 11**00** 1111 1111
  - b4 1 - 11 - 1**0**11 - 111**0** **0**111 - 1111 11**00** **0**111 1111

result: better clusters
(Hamming distance from 8 to 4)

The Visualization Challenge



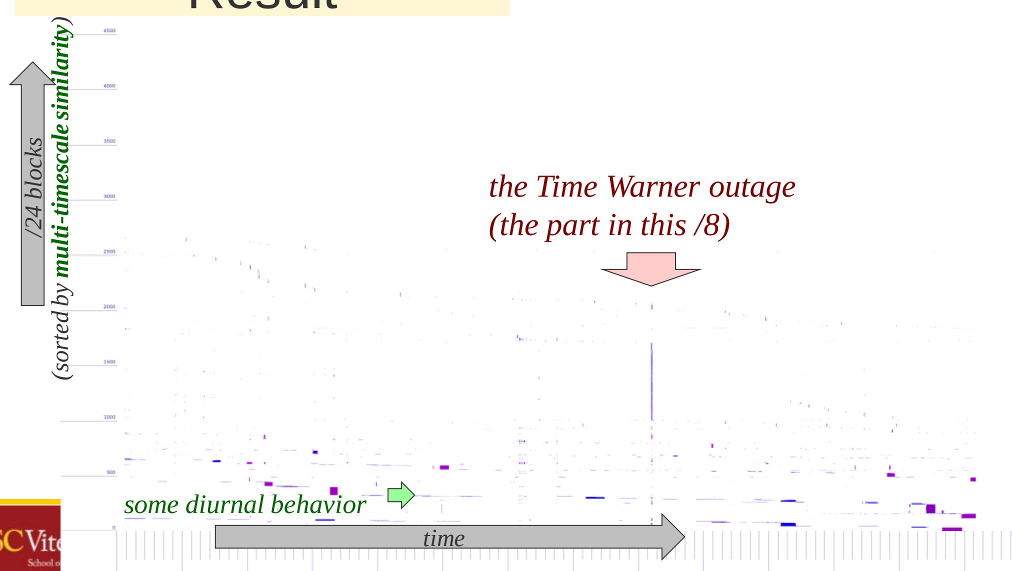
here ~1/4th (downsampled to fit the screen)
of 1/224th of the space (one /8 of IPv4)
and 1/12th of the duration (one quarter of ~3 years)
...what's happening? what trends? what's new?

USC Viterbi School of Engineering

90

One Visualization Result

here ~1/4th (downsampled to fit the screen)
of 1/224th of the space (one /8 of IPv4)
and 1/12th of the duration (one quarter of ~3 years)



USC Viterbi School of Engineering

91

Clustering to Discovery Dependencies

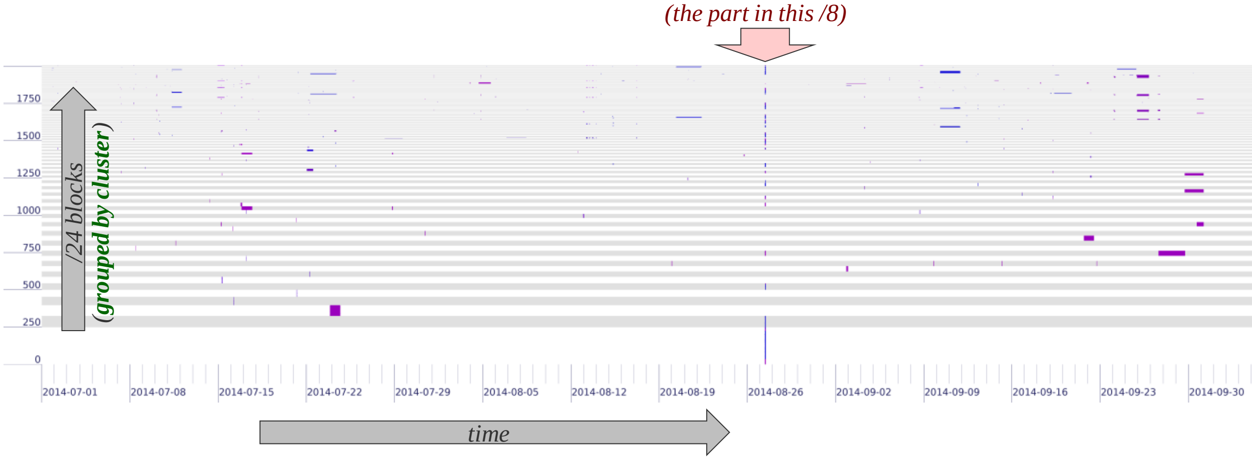
- visualization is nice, but humans can’t look at everything
- new clustering algorithms can *discovery dependencies*
  - insight: failure at the same time, multiple times => dependency
  - cluster on similarity of fail/recovery events

(Details: John Heidemann, Yuri Pradkin, and Aqib Nisar. Back Out: End-to-end Inference of Common Points-of-Failure in the Internet (extended). ISI-TR-724, February, 2018. https://www.isi.edu/%7ejohnh/PAPERS/Heidemann18b.html *.)*

One Clustering Result

1/224th of the space (one /8 of IPv4)
and 1/12th of the duration (one quarter of ~3 years)

the Time Warner outage (the part in this /8)



Iterative Clustering

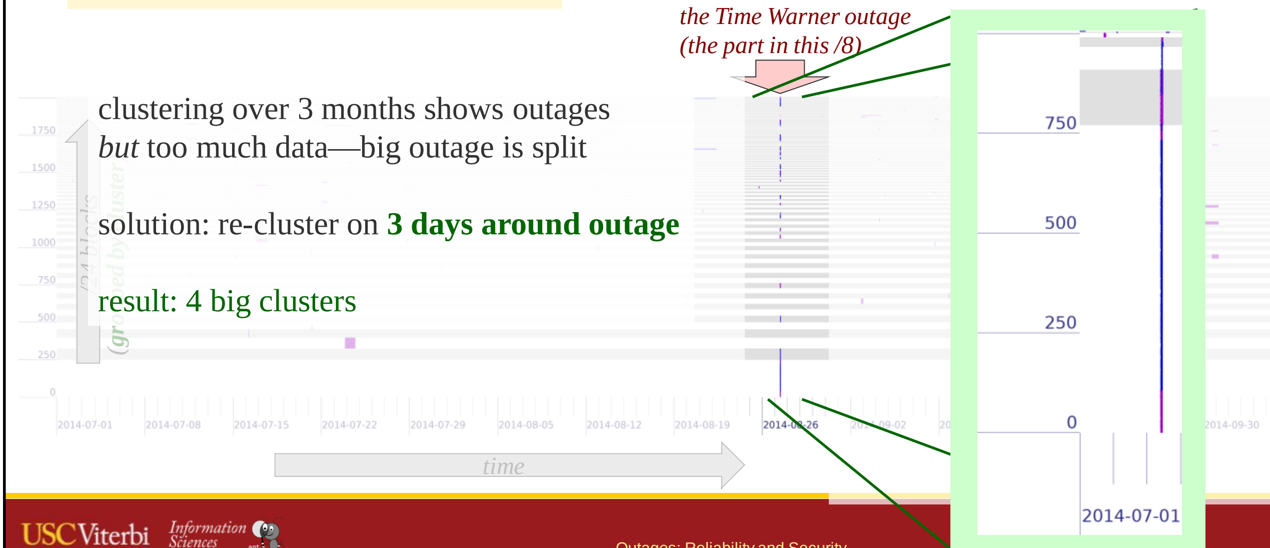
1/224th of the space (one /8 of IPv4)
and 1/12th of the duration (one quarter of ~3 years)
now just **3 days** of time

the Time Warner outage
(the part in this /8)

clustering over 3 months shows outages
but too much data—big outage is split

solution: re-cluster on **3 days around outage**

result: 4 big clusters



Clustering from Here

- just released clustering technical report
- from here…
  - does clustering relate to external information? (like power outages)
  - what are "normal" outages?
  - can we evaluate policy <=> reliablity?
  - what policy questions does this bring?

Next Steps

- can you use our data and approaches to improve Internet reliability and security?
- datasets: www.ImpactCyberTrust.org and https://ant.isi.edu/datasets/
  - data to inform policy?
  - historical data for long-term analysis?
  - compare to your new methods?
- code and papers: https://ant.isi.edu/