Observing the Global IPv4 Internet: What IP Addresses Show

John Heidemann

University of Southern California / Information Sciences Institute and CS Dept.

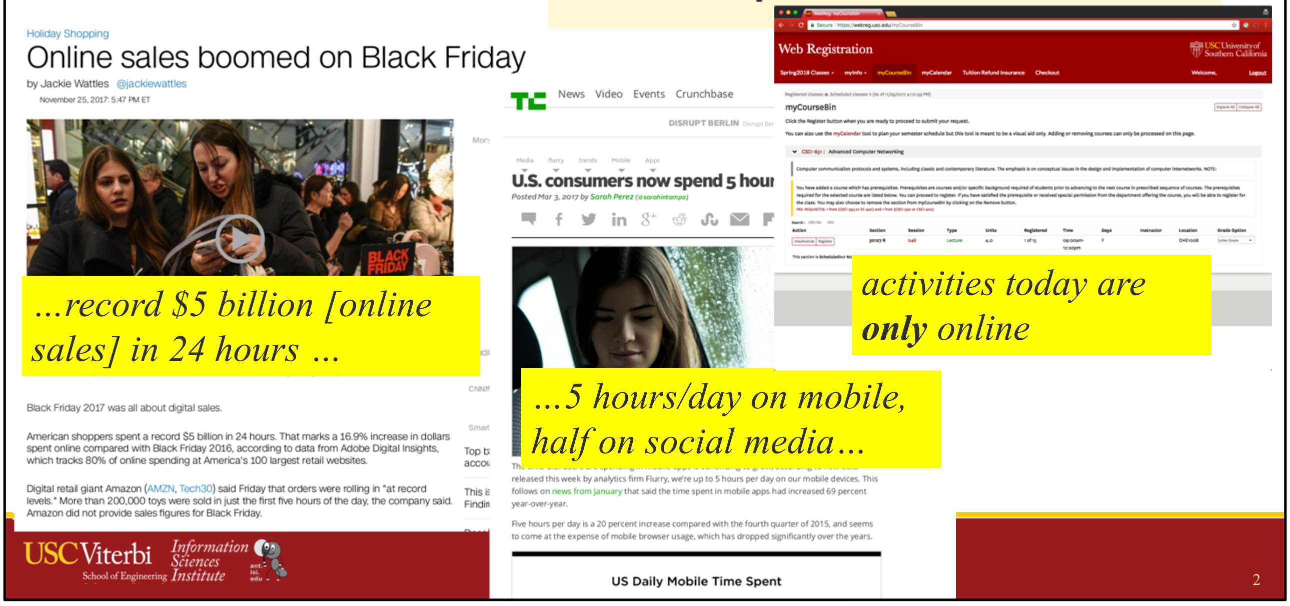
joint work with Guillermo Baltra, Asma Enayet, Yuri Pradkin, Xiao Song

prior contributors: Abdulla Alwabel, Ryan Bogutz, Aqib Nisar, Lin Quan

2021-06-17

This research is sponsored by NSF -2028279 (MINCEQ) and CNS-2007106 (EIEIO). Prior work was supported by NSF and DHS HSAPRRPA Cyber Security Division. The views herein are those of the authors and do not necessarily represent those of DHS or the U.S. Gov't.

The Internet is Important…



Holiday Shopping

Online sales boomed on Black Friday

by Jackie Wattles @jackiewattles

November 25, 2017: 5:47 PM ET

Black Friday 2017 was all about digital sales.

American shoppers spent a record $5 billion in 24 hours. That marks a 16.9% increase in dollars spent online compared with Black Friday 2016, according to data from Adobe Digital Insights, which tracks 80% of online spending at America's 100 largest retail websites.

Digital retail giant Amazon (AMZN, Tech30) said Friday that orders were rolling in "at record levels." More than 200,000 toys were sold in just the first five hours of the day, the company said. Amazon did not provide sales figures for Black Friday.

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

U.S. consumers now spend 5 hour

Posted Mar 3, 2017 by Sarah Perez (@sarahintampa)

released this week by analytics firm Flurry, we're up to 5 hours per day on our mobile devices. This follows on news from January that said the time spent in mobile apps had increased 69 percent year-over-year.

Five hours per day is a 20 percent increase compared with the fourth quarter of 2015, and seems to come at the expense of mobile browser usage, which has dropped significantly over the years.

US Daily Mobile Time Spent

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

Web Registration

myCourseBin

Click the Register button when you are ready to proceed to submit your request.

You can also use the myCalendar tool to plan your semester schedule but this tool is meant to be a visual aid only. Adding or removing courses can only be processed on this page.

CSCI-651 : Advanced Computer Networking

Computer communication protocols and systems, including classic and contemporary literature. The emphasis is on conceptual issues in the design and implementation of computer internetworks. NOTE:

You have added a course which has prerequisites. Prerequisites are courses and/or specific background required of students prior to advancing to the next course in prescribed sequence of courses. The prerequisites required for the selected course are listed below. You can proceed to register. If you have satisfied the prerequisite or received special permission from the department offering the course, you will be able to register for the class. You may also choose to remove the section from myCourseBin by clicking on the Remove button.

Action	Section	Session	Type	Units	Registered	Time	Days	Instructor	Location	Grade Option
Unschedule Register	30127 R	048	Lecture	4.0	1 of 15	09:00am-12:20pm	F		OHE1008	Letter Grade

activities today are ***only*** *online*

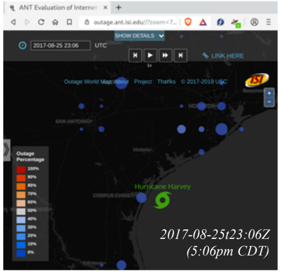
The *World* Is Important

hurricanes, floods, fires, blizzards…

Hurricane Harvey, August 2017

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

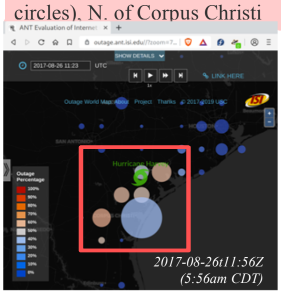
before landfall: **few outages**



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

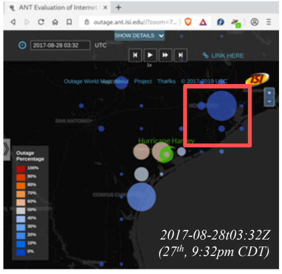
Harvey Landfall: 2017-08-26t03:00Z (10pm CDT)

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



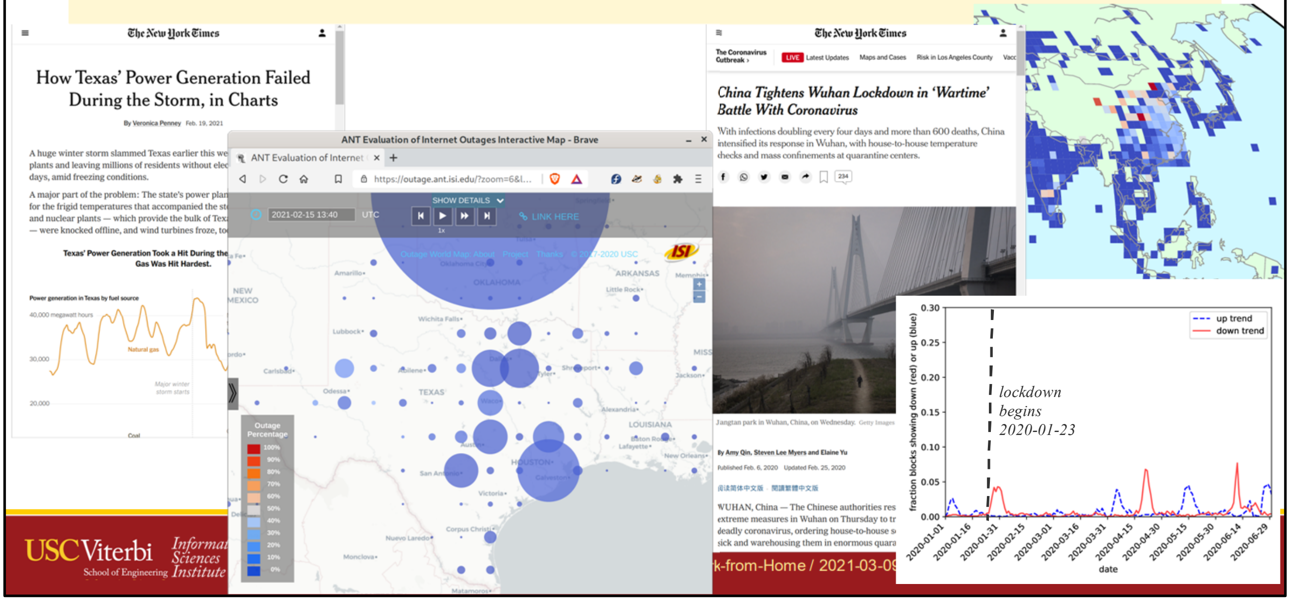
2017-08-26t11:56Z (5:56am CDT)

many outages (large circles), in Houston-flooding



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

Events are Changing the World



The New York Times

How Texas' Power Generation Failed During the Storm, in Charts

By Veronica Penney Feb. 19, 2021

A huge winter storm slammed Texas earlier this we plants and leaving millions of residents without elec days, amid freezing conditions.

A major part of the problem: The state's power plan for the frigid temperatures that accompanied the st and nuclear plants — which provide the bulk of Tex — were knocked offline, and wind turbines froze, to

Texas' Power Generation Took a Hit During the Gas Was Hit Hardest.

Power generation in Texas by fuel source
40,000 megawatt hours
30,000
20,000
Natural gas
Major winter storm starts
Coal

ANT Evaluation of Internet Outages Interactive Map - Brave

The New York Times

The Coronavirus Outbreak › LIVE Latest Updates Maps and Cases Risk in Los Angeles County Vacc

China Tightens Wuhan Lockdown in 'Wartime' Battle With Coronavirus

With infections doubling every four days and more than 600 deaths, China intensified its response in Wuhan, with house-to-house temperature checks and mass confinements at quarantine centers.

Jangtan park in Wuhan, China, on Wednesday. Getty Images

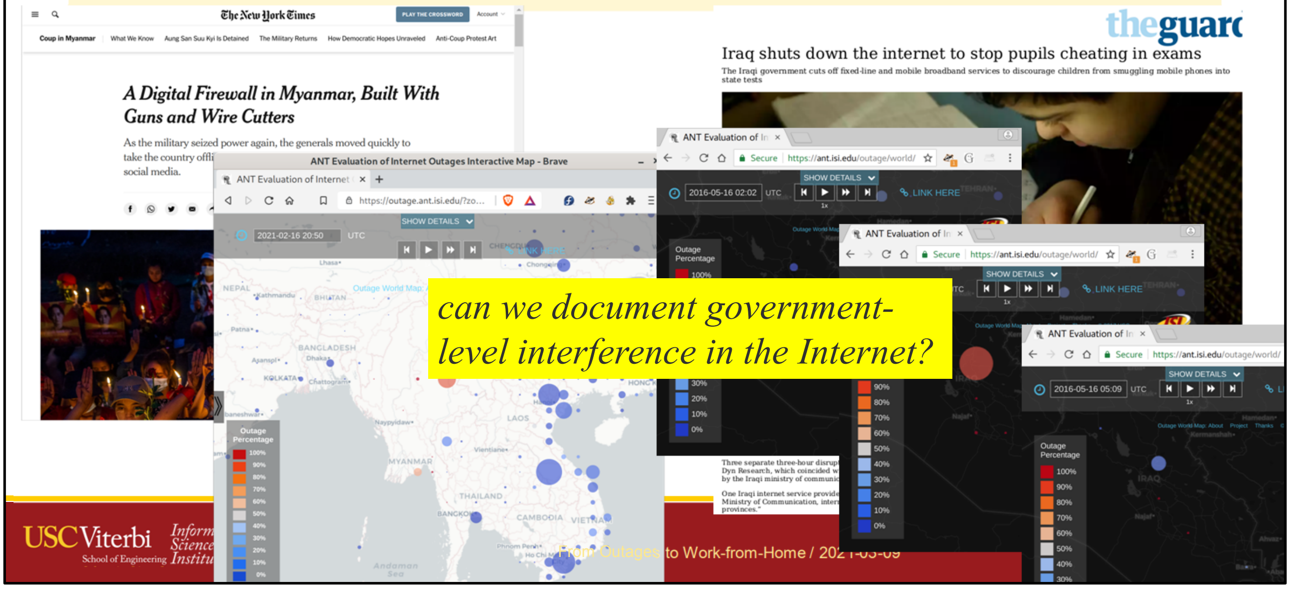
By Amy Qin, Steven Lee Myers and Elaine Yu

Published Feb. 6, 2020 Updated Feb. 25, 2020

阅读简体中文版 · 閱讀繁體中文版

WUHAN, China — The Chinese authorities res extreme measures in Wuhan on Thursday to tr deadly coronavirus, ordering house-to-house s sick and warehousing them in enormous quara

Countries Are Changing the World



A Digital Firewall in Myanmar, Built With Guns and Wire Cutters

As the military seized power again, the generals moved quickly to take the country offli… social media.

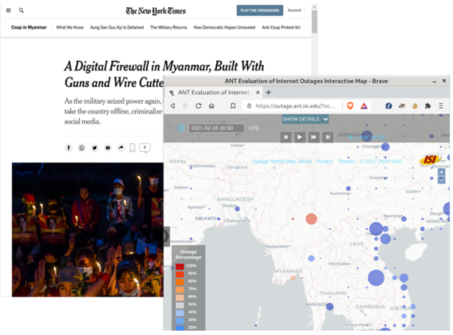
Iraq shuts down the internet to stop pupils cheating in exams

The Iraqi government cuts off fixed-line and mobile broadband services to discourage children from smuggling mobile phones into state tests

can we document government-level interference in the Internet?

Network Reliability Matters *Now*

in the Internet, in the world, and how they connect…



communication without **intentional network interference**

Harvey Landfall: 2017-08-26t03:00Z (10pm CDT)

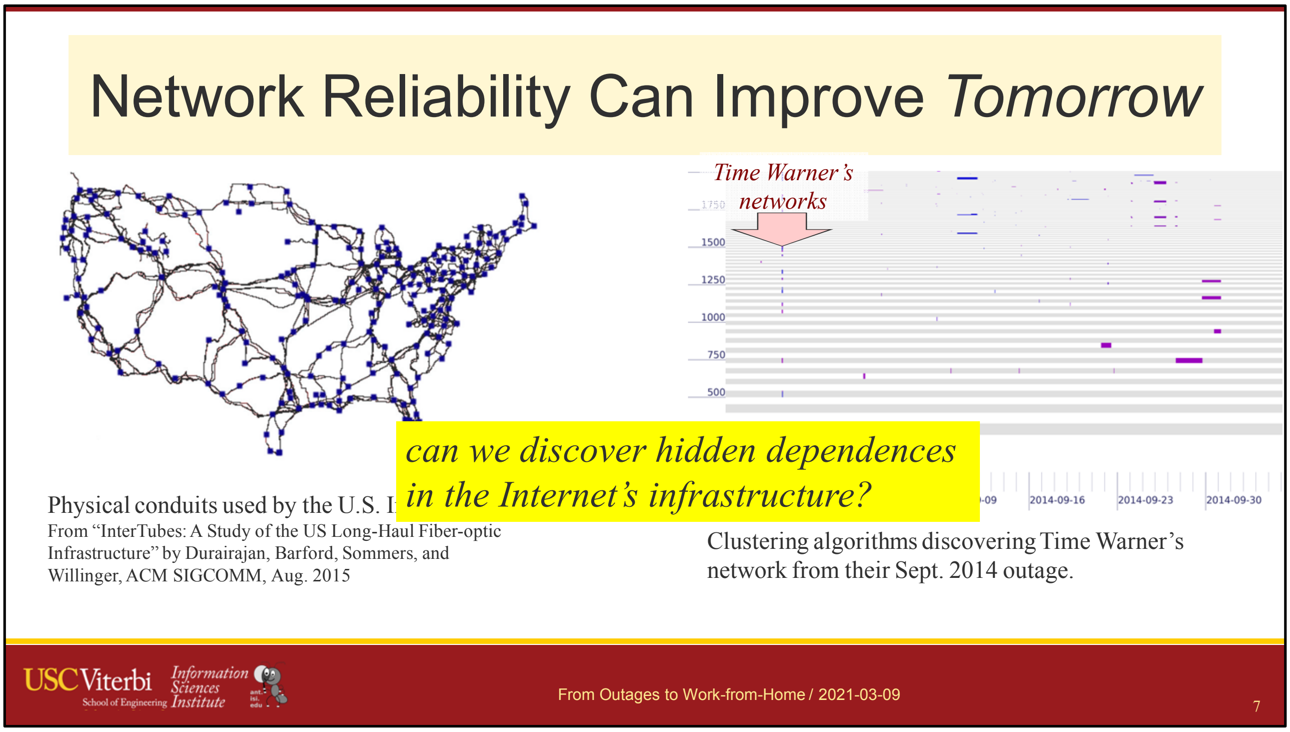
can we provide near-real-time results to help response?

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

2017-08-26 11:23 UTC
SHOW DETAILS
LINK HERE
Hurricane Harvey
Outage Percentage

speedy **physical recovery** to natural disasters

CDNs with **choices** where to serve customers



Network Reliability Can Improve *Tomorrow*

Time Warner's networks

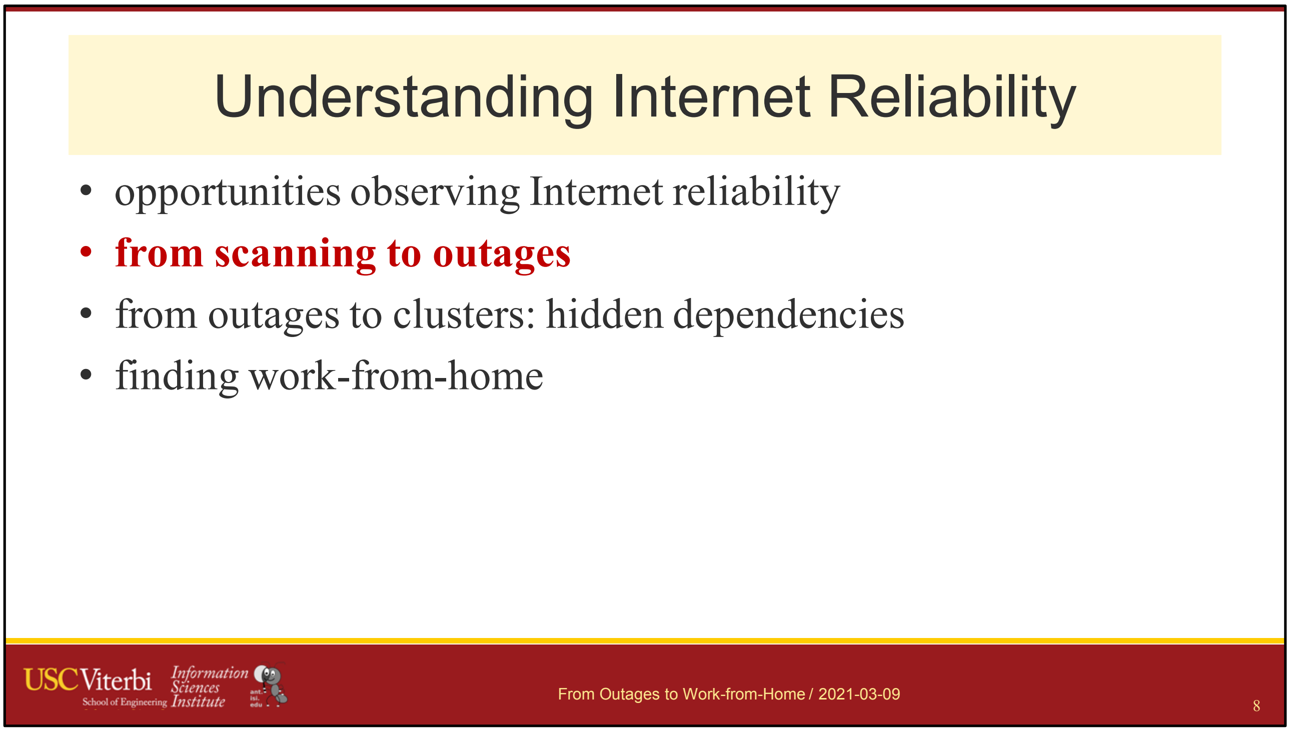
1750
1500
1250
1000
750
500
-09
2014-09-16
2014-09-23
2014-09-30

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

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

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



Understanding Internet Reliability

- opportunities observing Internet reliability
- **from scanning to outages**
- from outages to clusters: hidden dependencies
- finding work-from-home

The IPv4 Internet

we scan the IPv4 Internet (since 2006!)

2^{32} addresses (~4 billion)

usually written: 4 parts, each 8-bits

192.0.2.1 (from 0.0.0.0 to 255.255.255.255)

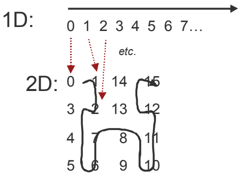
address **blocks:** adjacent addresses with same first *n* bits

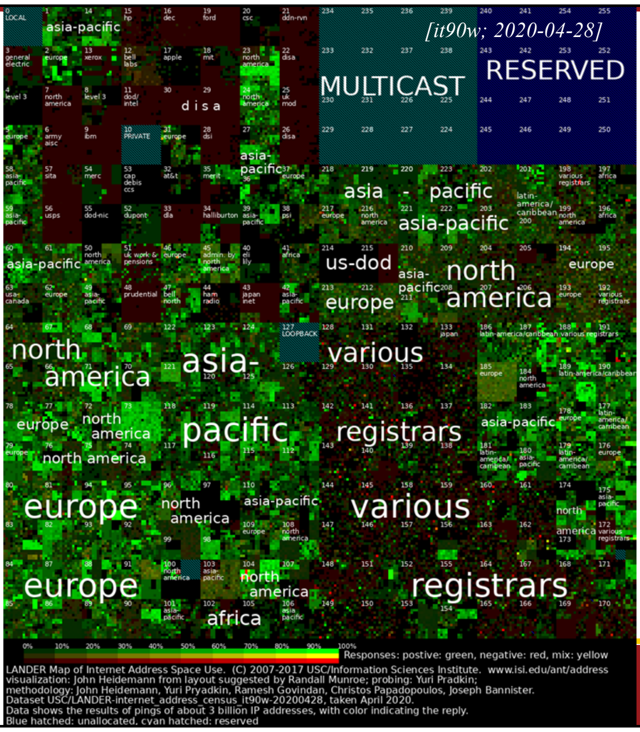
192.0.*.* /16

or just 192.0/16

(prefix=192.0, n=16)

squares on the map





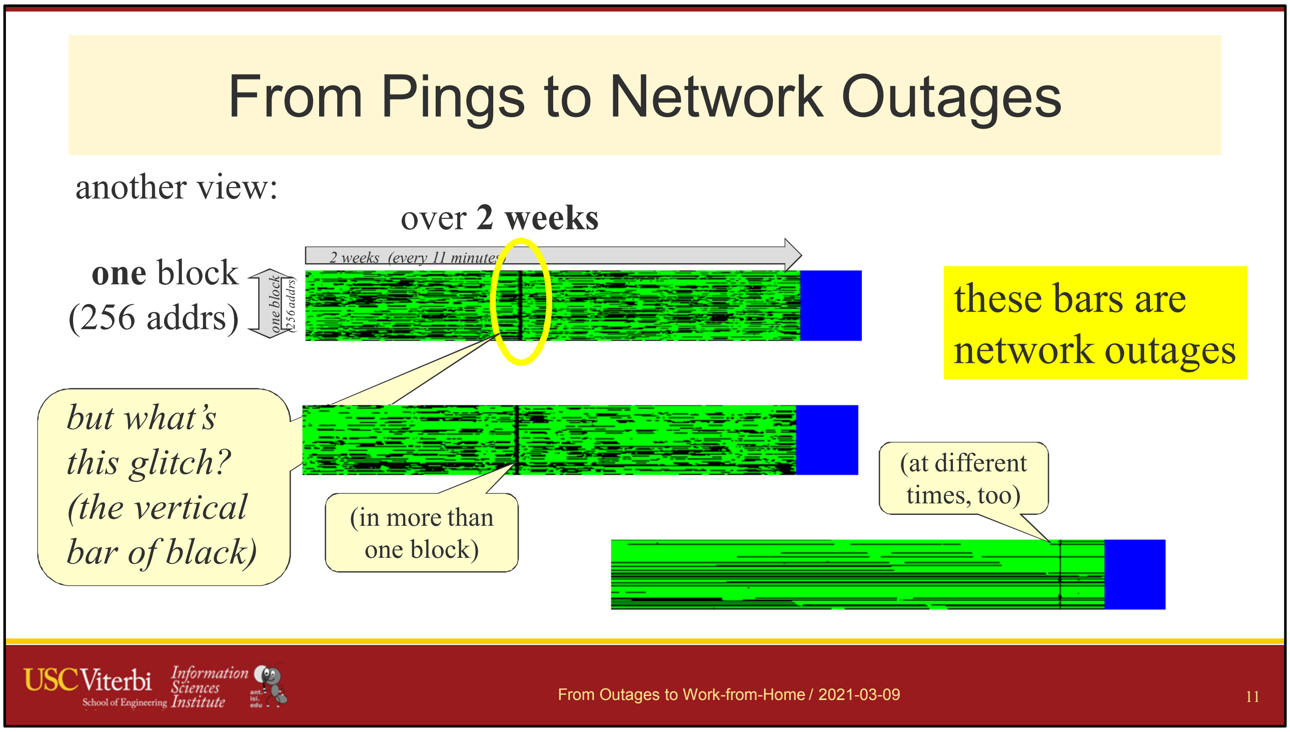
Responses: postive: green, negative: red, mix: yellow

LANDER Map of Internet Address Space Use. (C) 2007-2017 USC/Information Sciences Institute. www.isi.edu/ant/address
visualization: John Heidemann from layout suggested by Randall Munroe; probing: Yuri Pradkin;
methodology: John Heidemann, Yuri Pryadkin, Ramesh Govindan, Christos Papadopoulos, Joseph Bannister.
Dataset USC/LANDER-internet_address_census_it90w-20200428, taken April 2020.
Data shows the results of pings of about 3 billion IP addresses, with color indicating the reply.
Blue hatched: unallocated, cyan hatched: reserved

[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.
- this data is from 2011



From Pings to Network Outages

another view:

over **2 weeks**

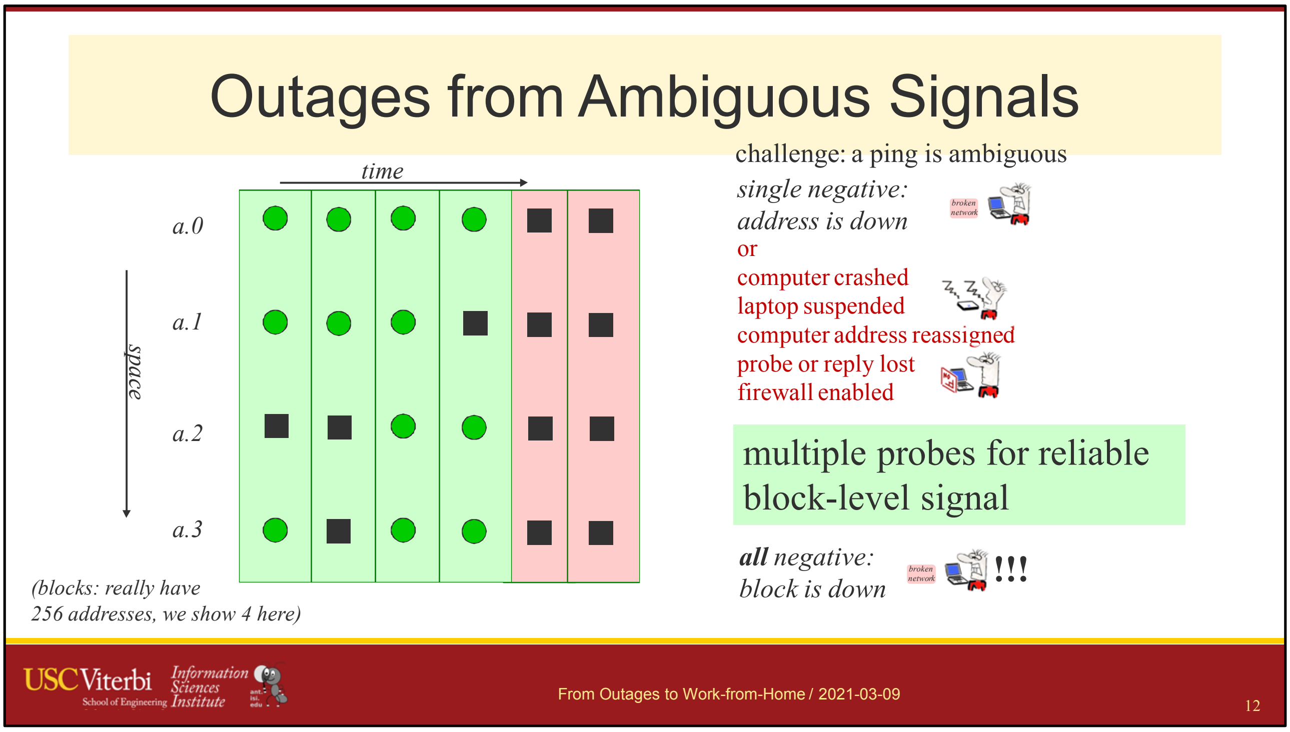
one block
(256 addrs)

these bars are network outages

but what's this glitch? (the vertical bar of black)

(in more than one block)

(at different times, too)



Outages from Ambiguous Signals

a.0	●	●	●	●	■	■
a.1	●	●	●	■	■	■
a.2	■	■	●	●	■	■
a.3	●	■	●	●	■	■

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

challenge: a ping is ambiguous

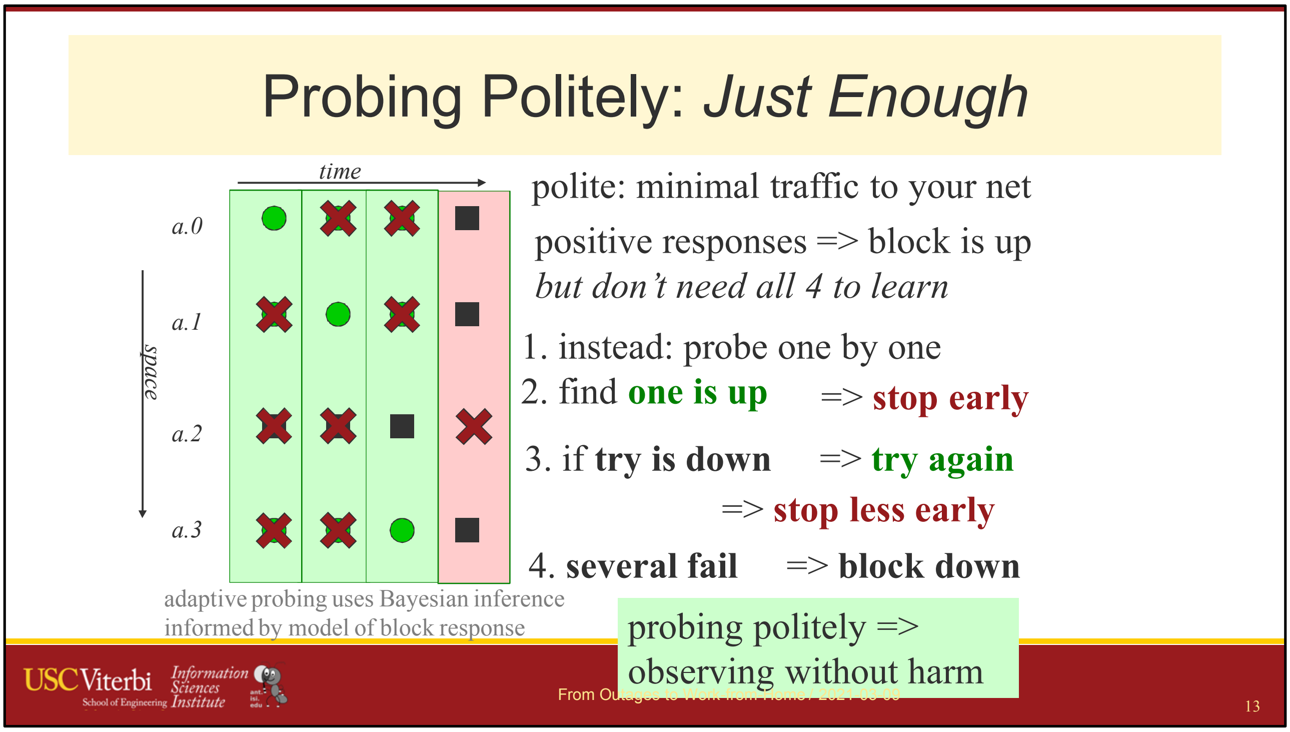
single negative: address is down

or

computer crashed
laptop suspended
computer address reassigned
probe or reply lost
firewall enabled

multiple probes for reliable block-level signal

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

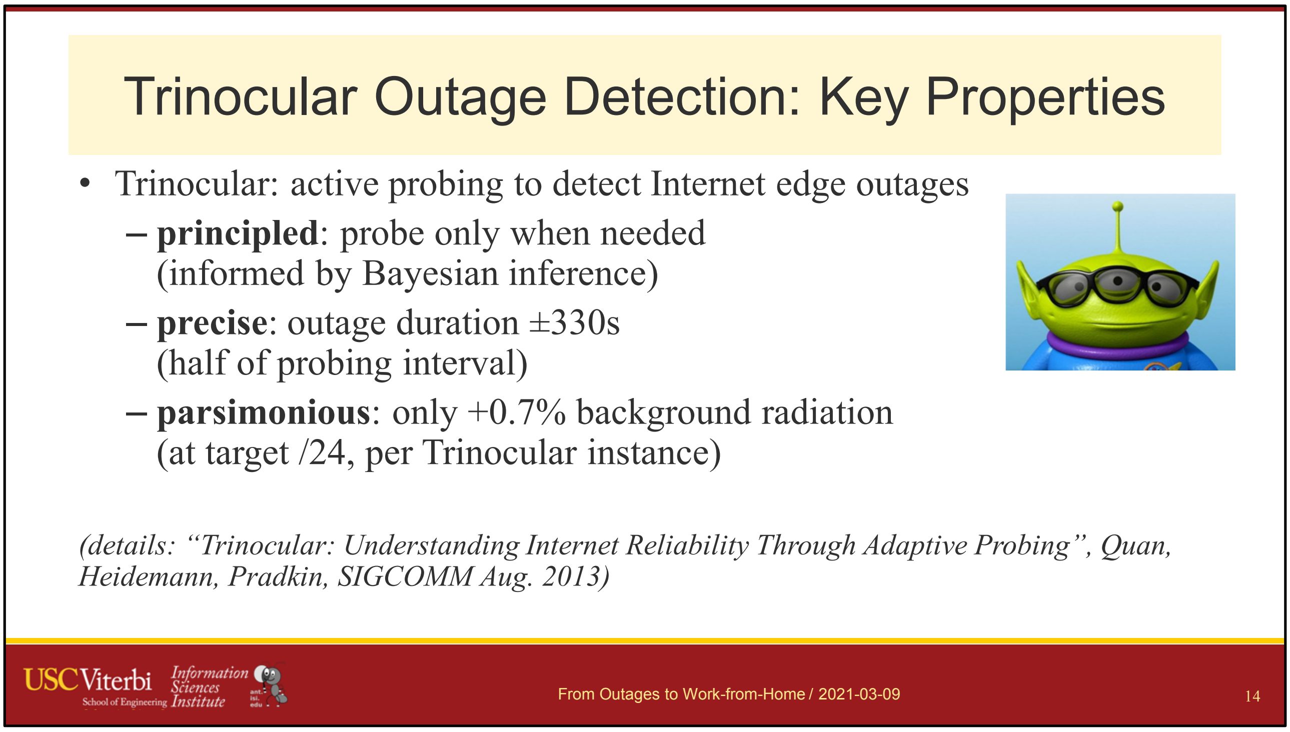
1. instead: probe one by one
2. find **one is up** => **stop early**
3. if **try is down** => **try again**
=> **stop less early**
4. **several fail** => **block down**

adaptive probing uses Bayesian inference
informed by model of block response

probing politely =>
observing without harm

From Outages to Work-from-Home / 2021-03-09

13



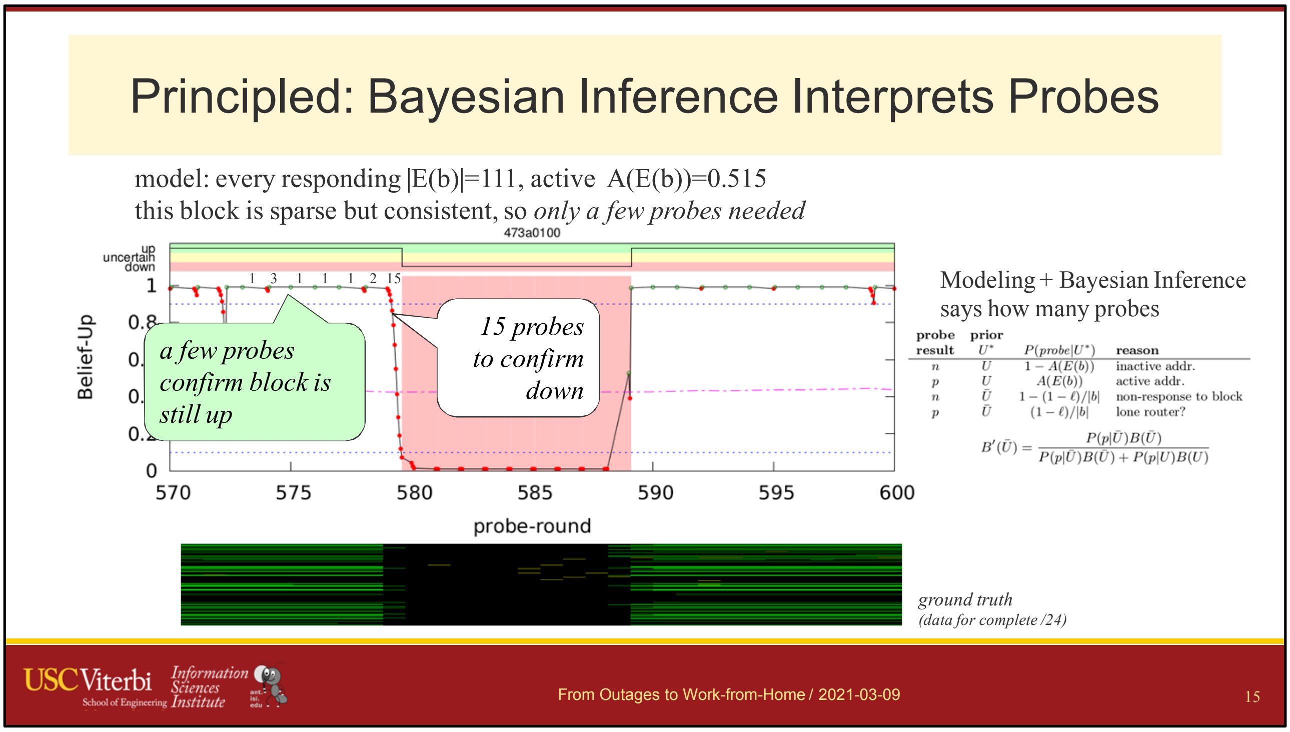
Trinocular Outage Detection: Key Properties

- 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)

From Outages to Work-from-Home / 2021-03-09

14



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*

a few probes confirm block is still up

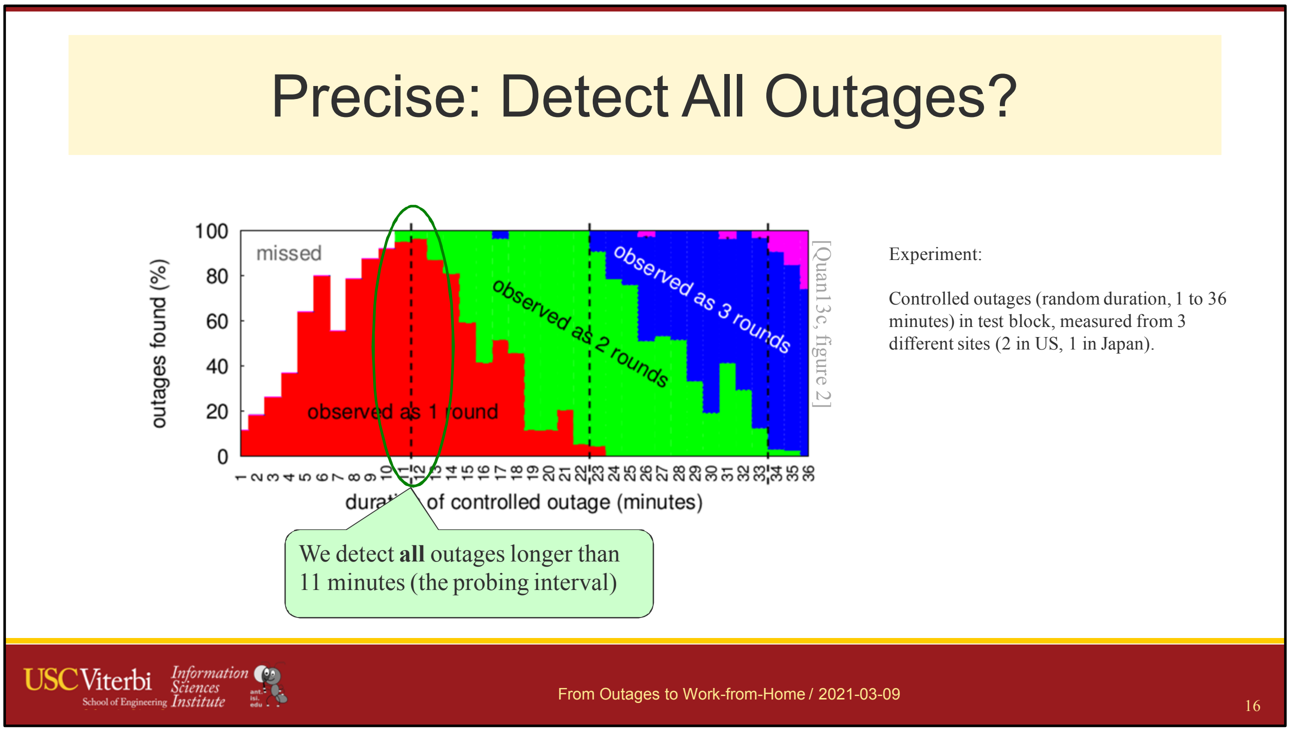
15 probes to confirm down

Modeling + Bayesian Inference says how many probes

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

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

ground truth
(data for complete /24)



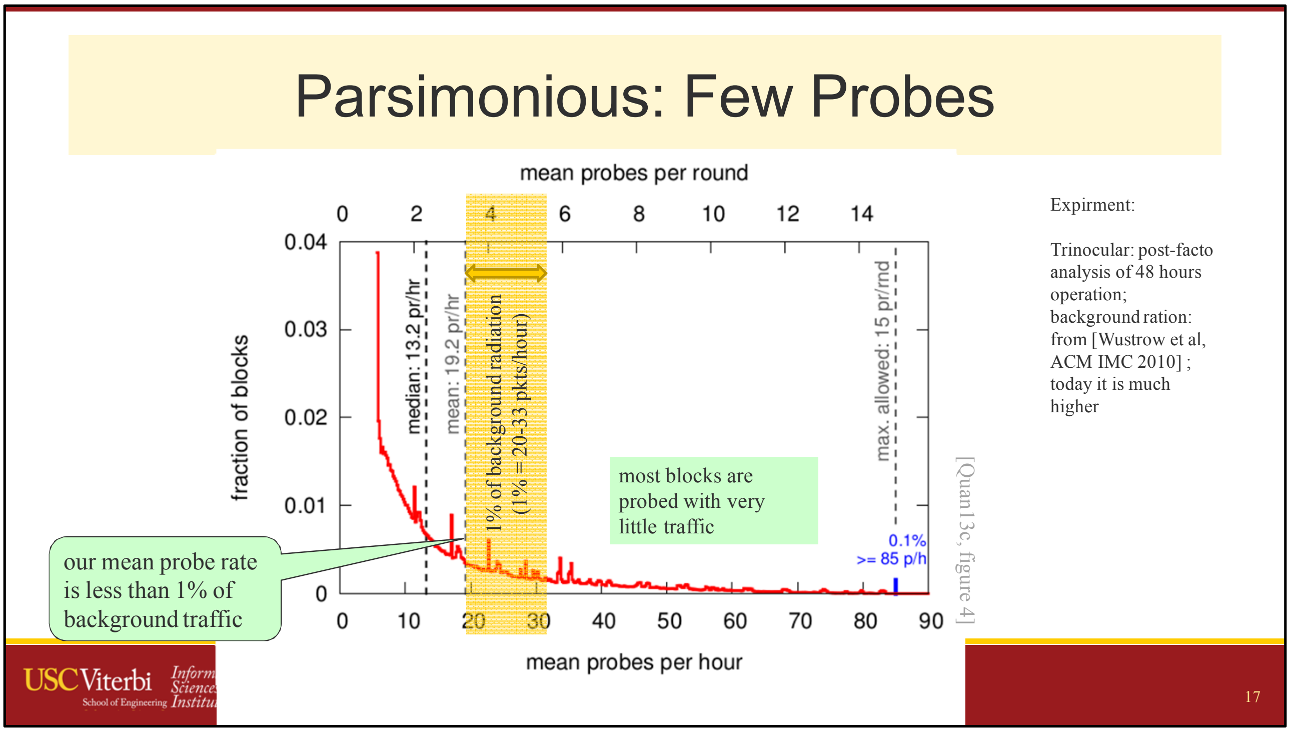
Precise: Detect All Outages?

[Quan13c, figure 2]

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: Few Probes

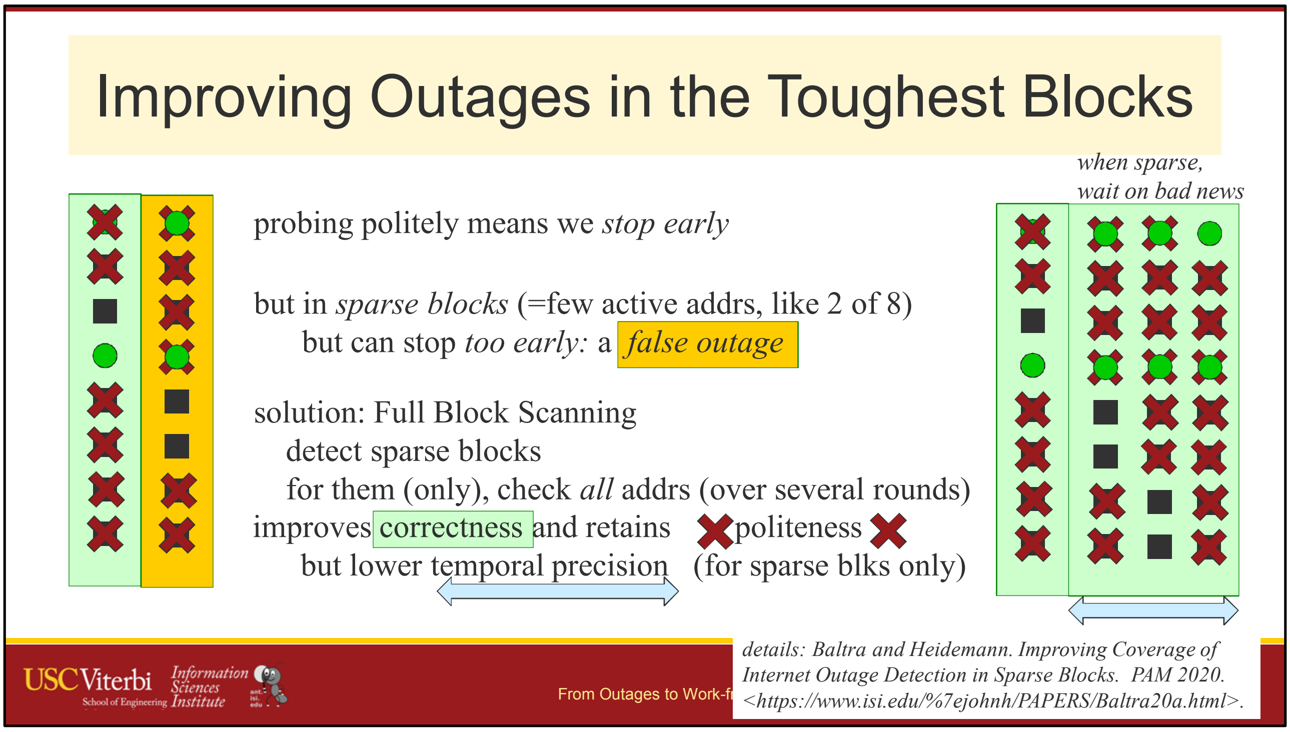
most blocks are probed with very little traffic

our mean probe rate is less than 1% of background traffic

Expriment:

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

17



Improving Outages in the Toughest Blocks

probing politely means we *stop early*

but in *sparse blocks* (=few active addrs, like 2 of 8)
but can stop *too early:* a *false outage*

solution: Full Block Scanning
detect sparse blocks
for them (only), check *all* addrs (over several rounds)
improves correctness and retains ✗politeness ✗
but lower temporal precision (for sparse blks only)

when sparse,
wait on bad news

From Outages to Work-f

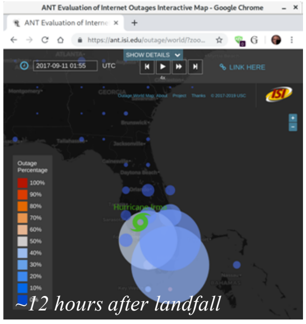
details: Baltra and Heidemann. Improving Coverage of Internet Outage Detection in Sparse Blocks. PAM 2020. <https://www.isi.edu/%7ejohnh/PAPERS/Baltra20a.html>.

Impact of Outage Detection

- quantified impact of hurricanes
  - previously: Harvey (2017)
  - next: Irma (2017)
- outages in operational networks
- near-real time reporting

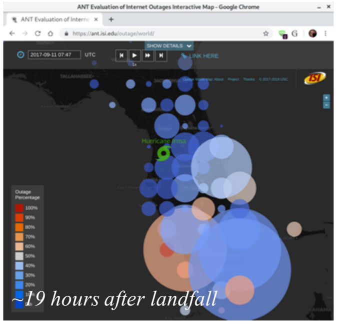
Hurricane Irma: Watching Recovery

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


ANT Evaluation of Internet Outages Interactive Map - Google Chrome
https://ant.isi.edu/outage/world/?zoo...
SHOW DETAILS
2017-09-11 01:55
UTC
LINK HERE
Hurricane Irma
Outage Percentage
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%

~12 hours after landfall

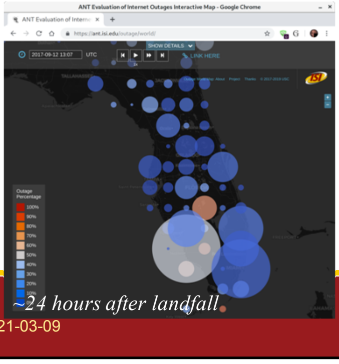
Irma landfall: 2017-09-10t13:10Z at Cudjoe Key, Florida



~19 hours after landfall

(play)

https://ant.isi.edu/url/irma2017/

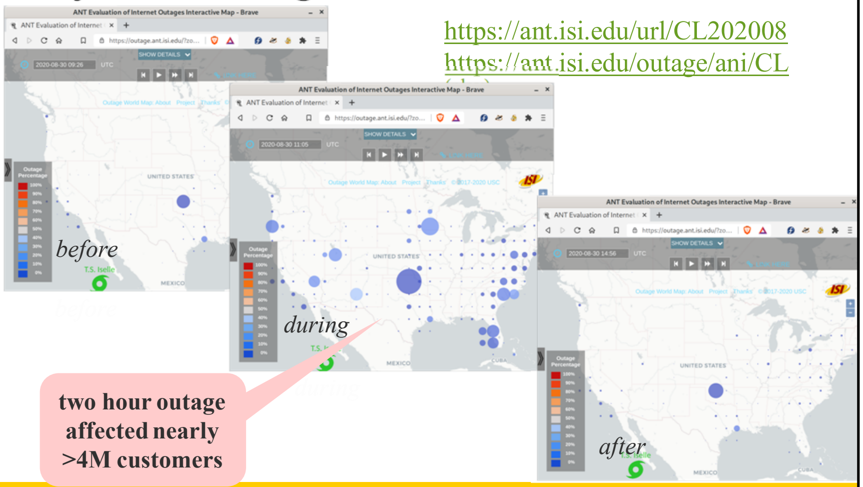

ANT Evaluation of Internet Outages Interactive Map - Google Chrome
SHOW DETAILS
2017-09-12 13:07
UTC
LINK HERE
Outage Percentage

~24 hours after landfall

Outages in Operational Networks: CenturyLink, August 2020

we also see problems due to network ops

- this dataset:
  - 5M blocks
  - all of 2020q3
- events:
  - CenturyLink outage on 2020-08-30 starting 9:55Z
  - >4 million customers



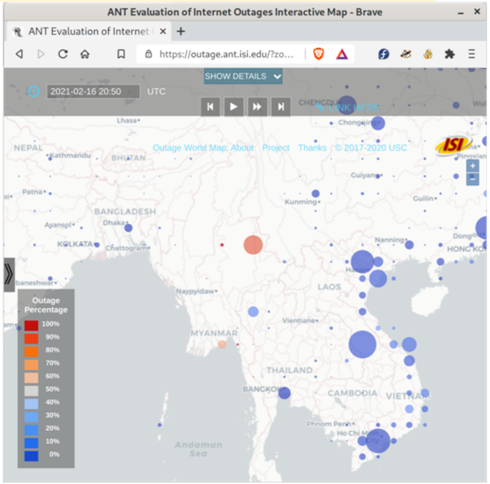
https://ant.isi.edu/url/CL202008
https://ant.isi.edu/outage/ani/CL

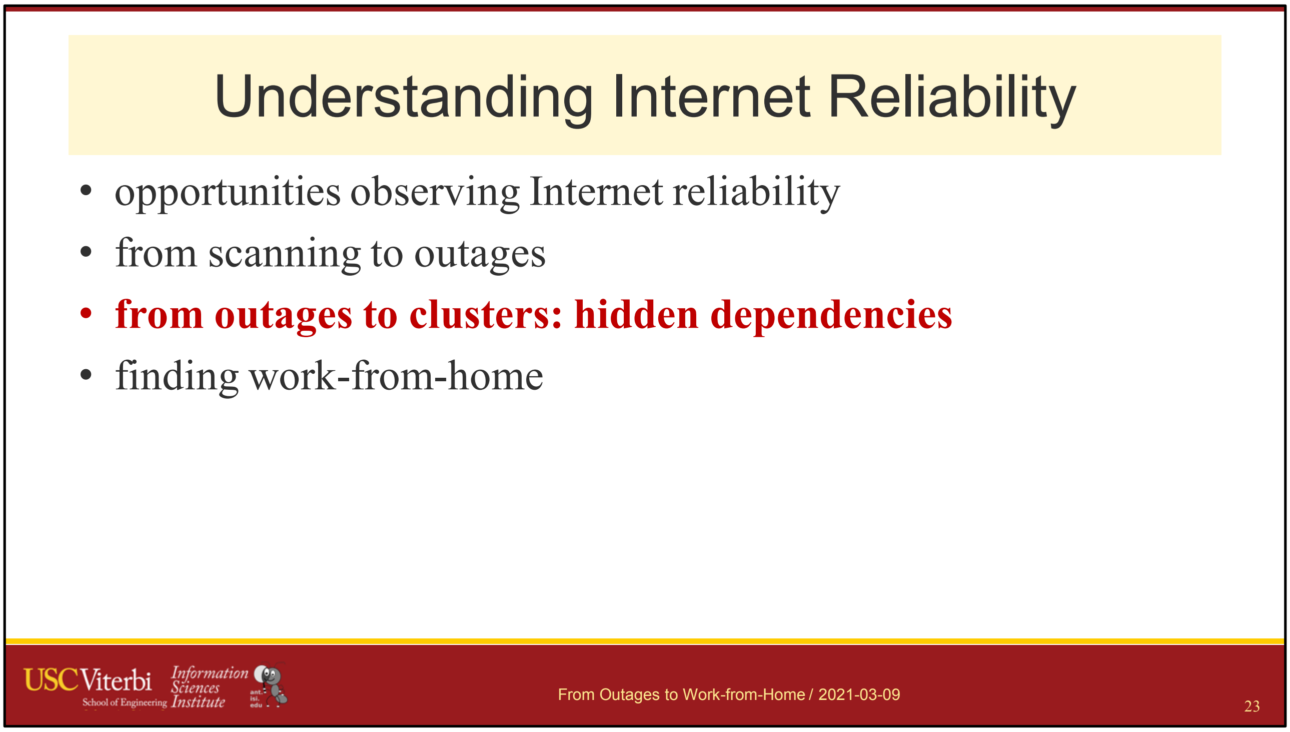
two hour outage affected nearly >4M customers

Near-Real Time Reporting (Now!)

- https://outage.ant.isi.edu/
- outages 24x7, within ~2h of observation
- visualized in your browser
  - circle size: *number* of blocks out
  - color: *percent* of blocks out
  - pan in geography and time
- goals:
  - support first responders
  - support the general public
  - global coverage

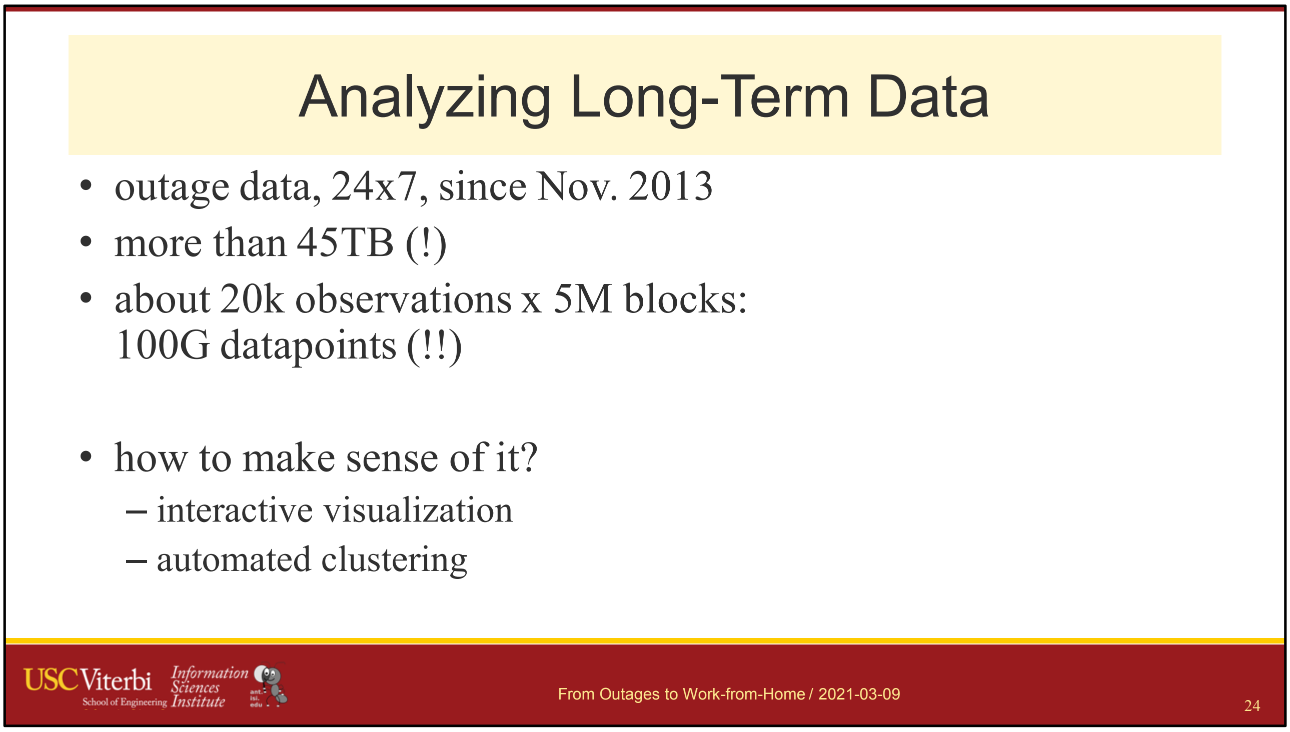
Myanmar, Internet shutdown 2021-02-16, 2 weeks after a military coup
https://ant.isi.edu/url/mm210206


ANT Evaluation of Internet Outages Interactive Map - Brave
ANT Evaluation of Internet
https://outage.ant.isi.edu/?zo...
SHOW DETAILS
2021-02-16 20:50 UTC
Outage World Map About Project Thanks © 2017-2020 USC
Outage Percentage
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
NEPAL
BHUTAN
BANGLADESH
MYANMAR
LAOS
THAILAND
CAMBODIA
VIETNAM
Andaman Sea



Understanding Internet Reliability

- opportunities observing Internet reliability
- from scanning to outages
- **from outages to clusters: hidden dependencies**
- finding work-from-home

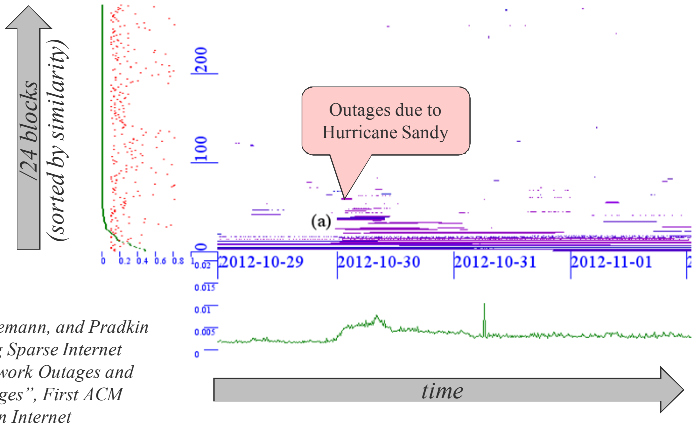


Analyzing Long-Term Data

- outage data, 24x7, since Nov. 2013
- more than 45TB (!)
- about 20k observations x 5M blocks:
  100G datapoints (!!)

- how to make sense of it?
  - interactive visualization
  - automated clustering

Non-Geographic Visualizations: the *Network* in Outages

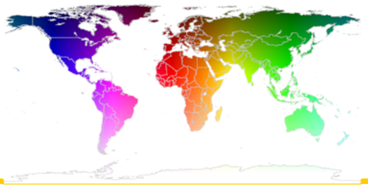


Outages due to Hurricane Sandy

goal: reveal patterns

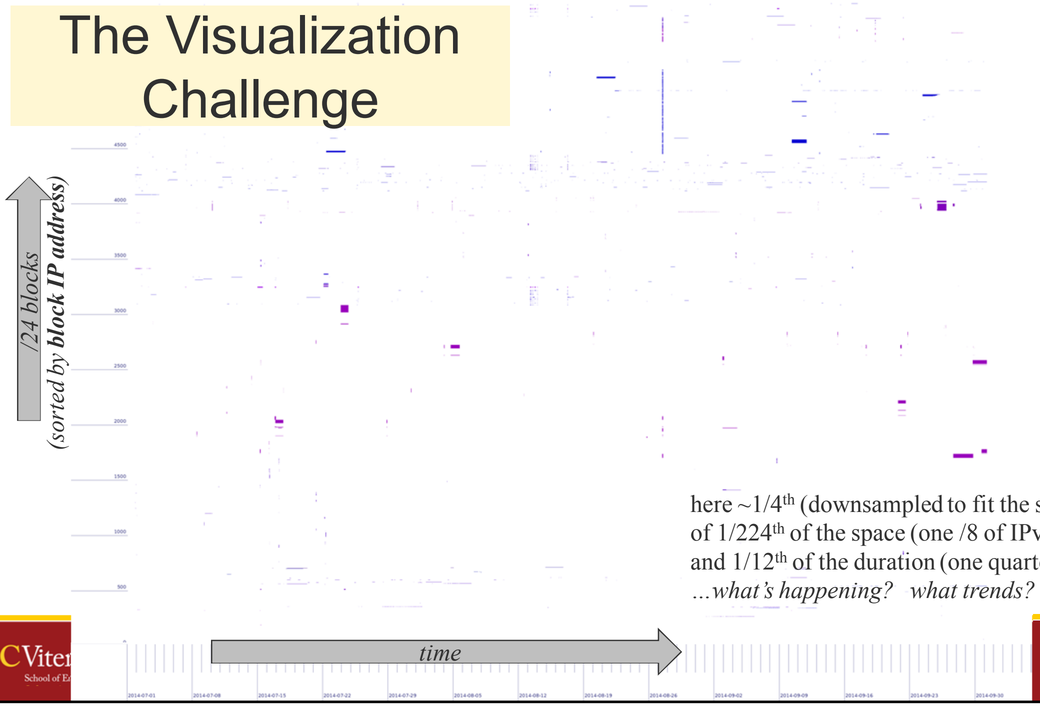
find dependencies among networks

(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

The Visualization Challenge

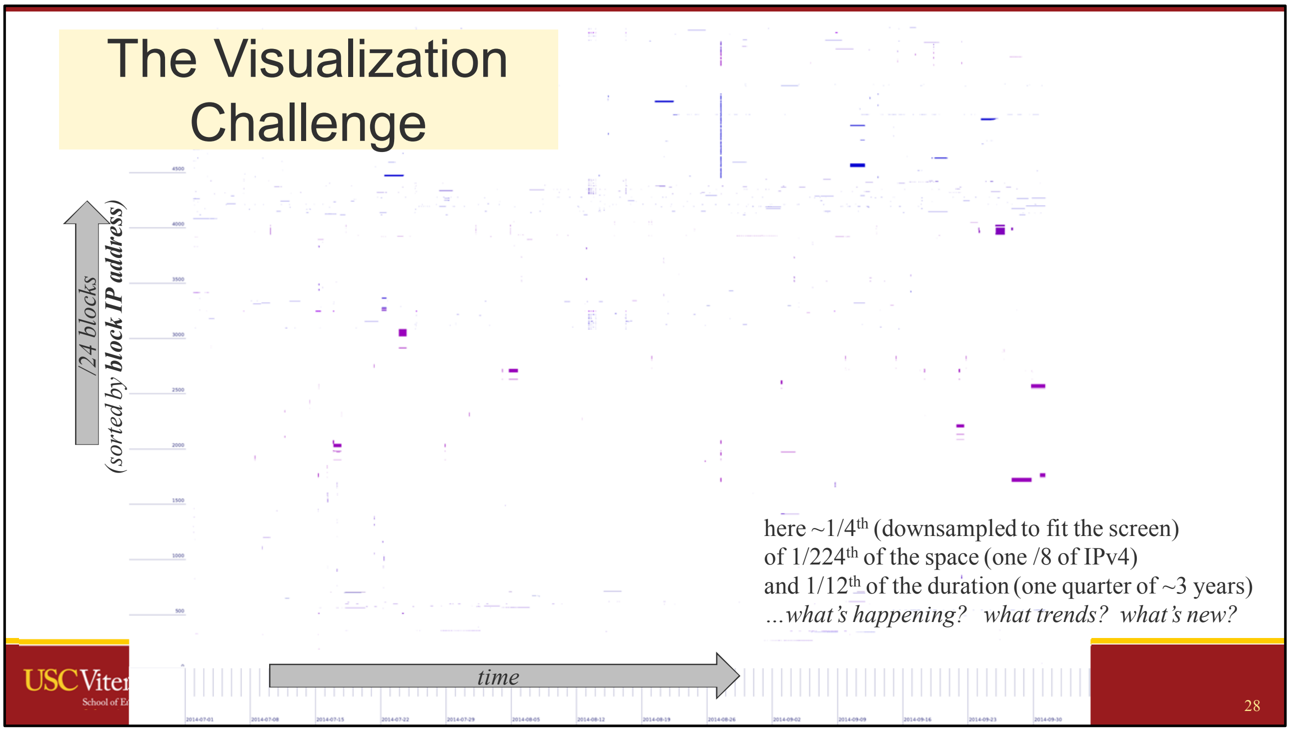


here ~1/4th (downsampled to fit the screen)
of 1/224th of the space (one /8 of IPv4)
and 1/12th of the duration (one quarter of ~3 years)
...what's happening? what trends? what's new?

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*

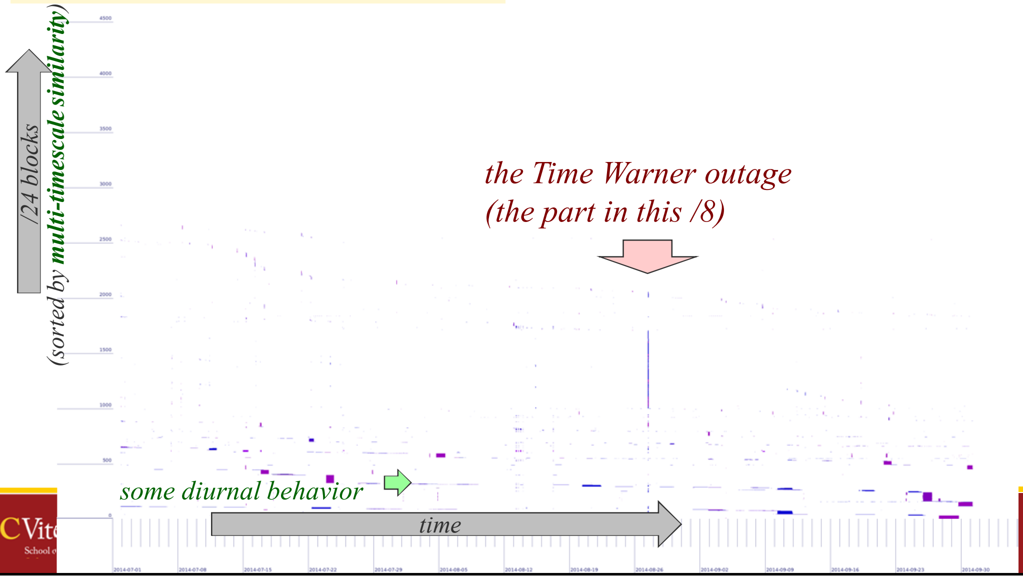


The Visualization Challenge

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?

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)



/24 blocks

(sorted by ***multi-timescale similarity****)*

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

some diurnal behavior

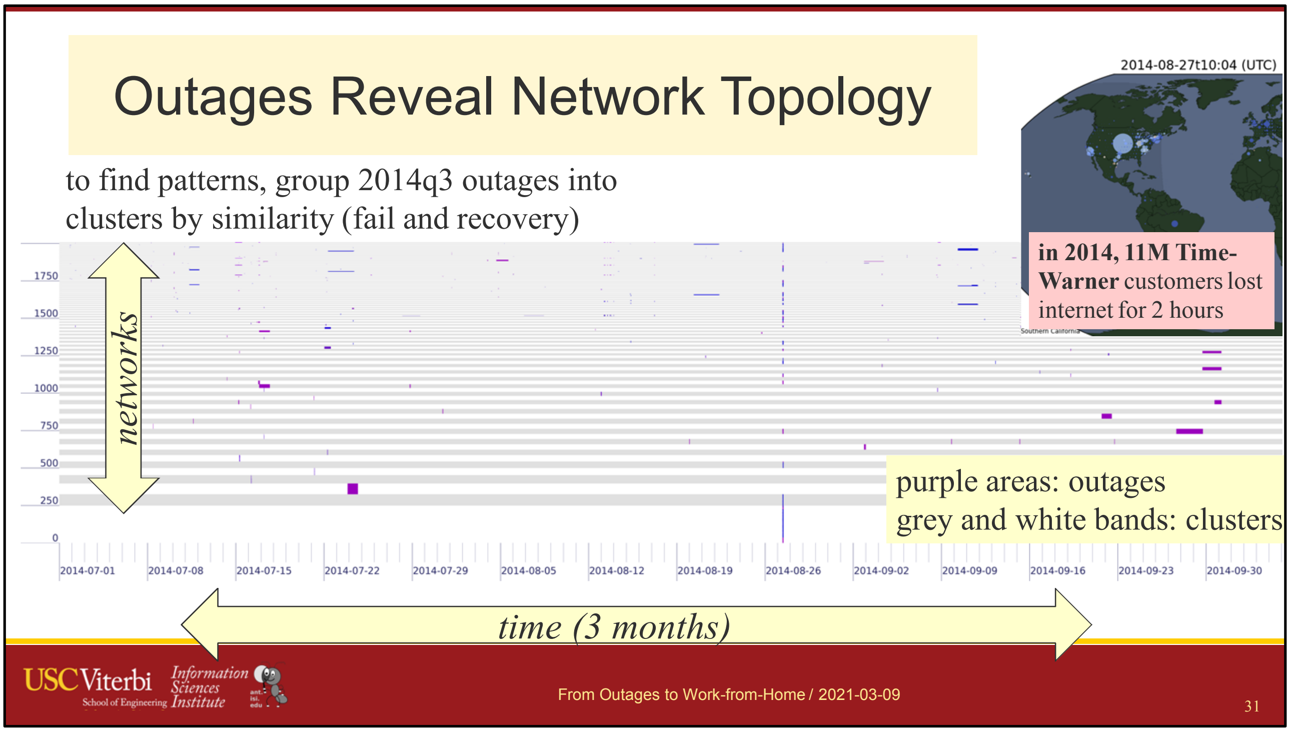
time

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 .)



Outages Reveal Network Topology

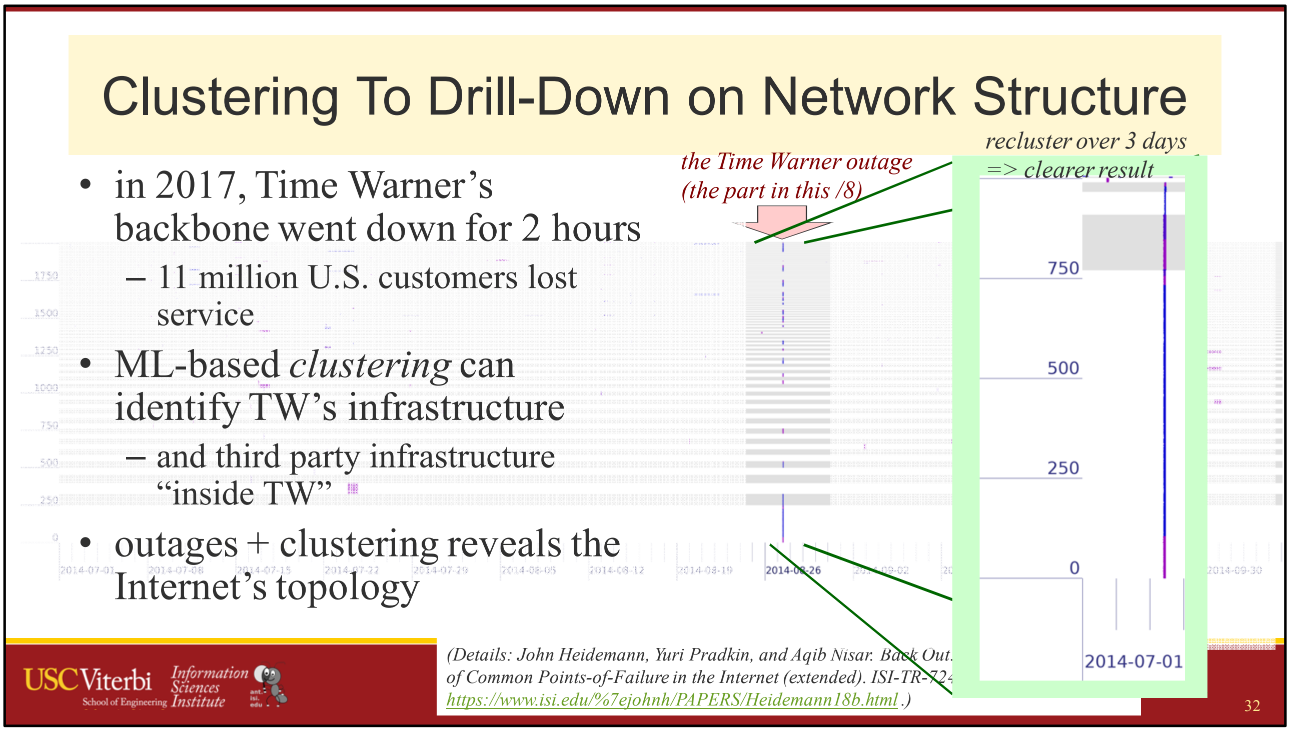
to find patterns, group 2014q3 outages into clusters by similarity (fail and recovery)

in 2014, 11M Time-Warner customers lost internet for 2 hours

purple areas: outages
grey and white bands: clusters

From Outages to Work-from-Home / 2021-03-09

31



Clustering To Drill-Down on Network Structure

- in 2017, Time Warner's backbone went down for 2 hours
  - 11 million U.S. customers lost service
- ML-based *clustering* can identify TW's infrastructure
  - and third party infrastructure "inside TW"
- outages + clustering reveals the Internet's topology

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

recluster over 3 days => clearer result

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

32

Understanding Internet Reliability

- opportunities observing Internet reliability
- from scanning to outages
- from outages to clusters: hidden dependencies
- **finding work-from-home**

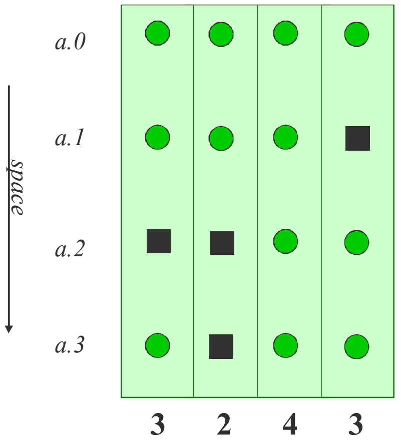
Q: Can We find Work-from-Home from Changes in IPv4 Address Usage?

Goal:

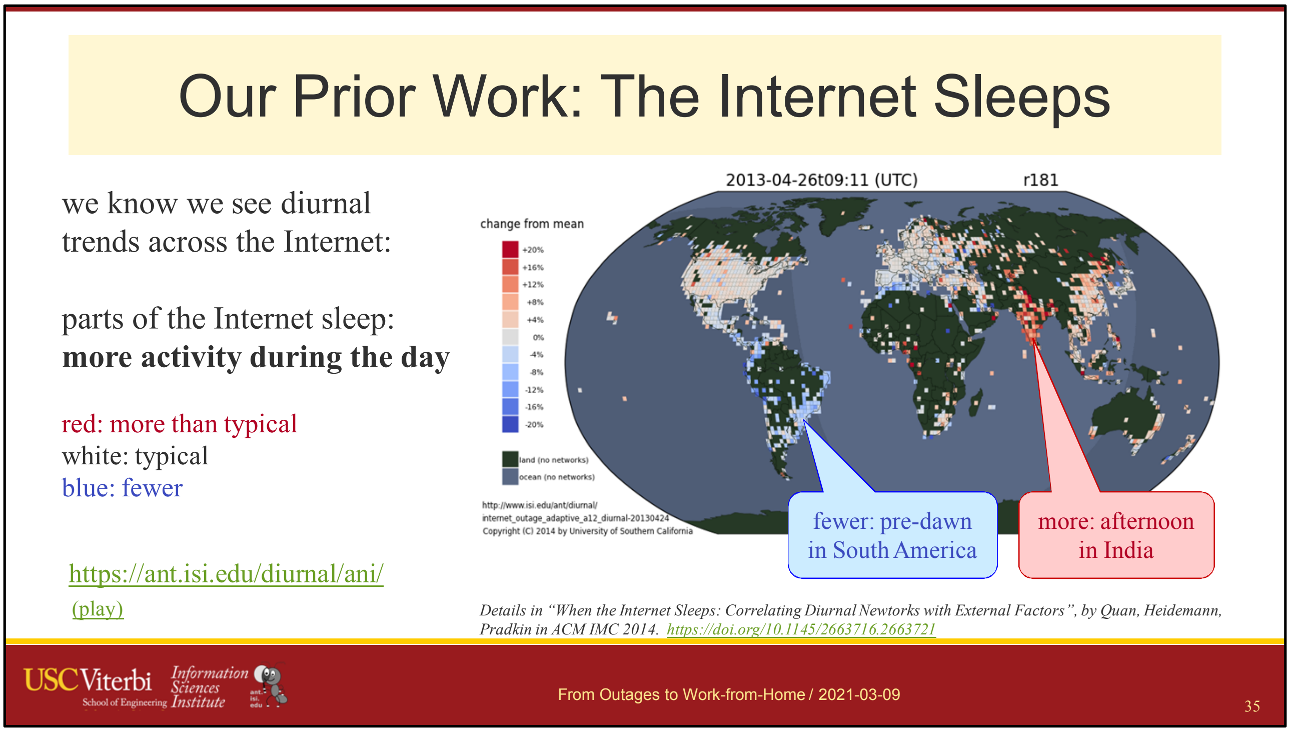
- do people *really* work-from home?
- can we confirm compliance?
- globally

Insight:

- when we probe all these addresses…
- we learn how the Internet "moves"
  - as computers are turned on and off
- so we learn how *people* move
  - as laptops come and go



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



Our Prior Work: The Internet Sleeps

we know we see diurnal trends across the Internet:

parts of the Internet sleep:
more activity during the day

red: more than typical
white: typical
blue: fewer

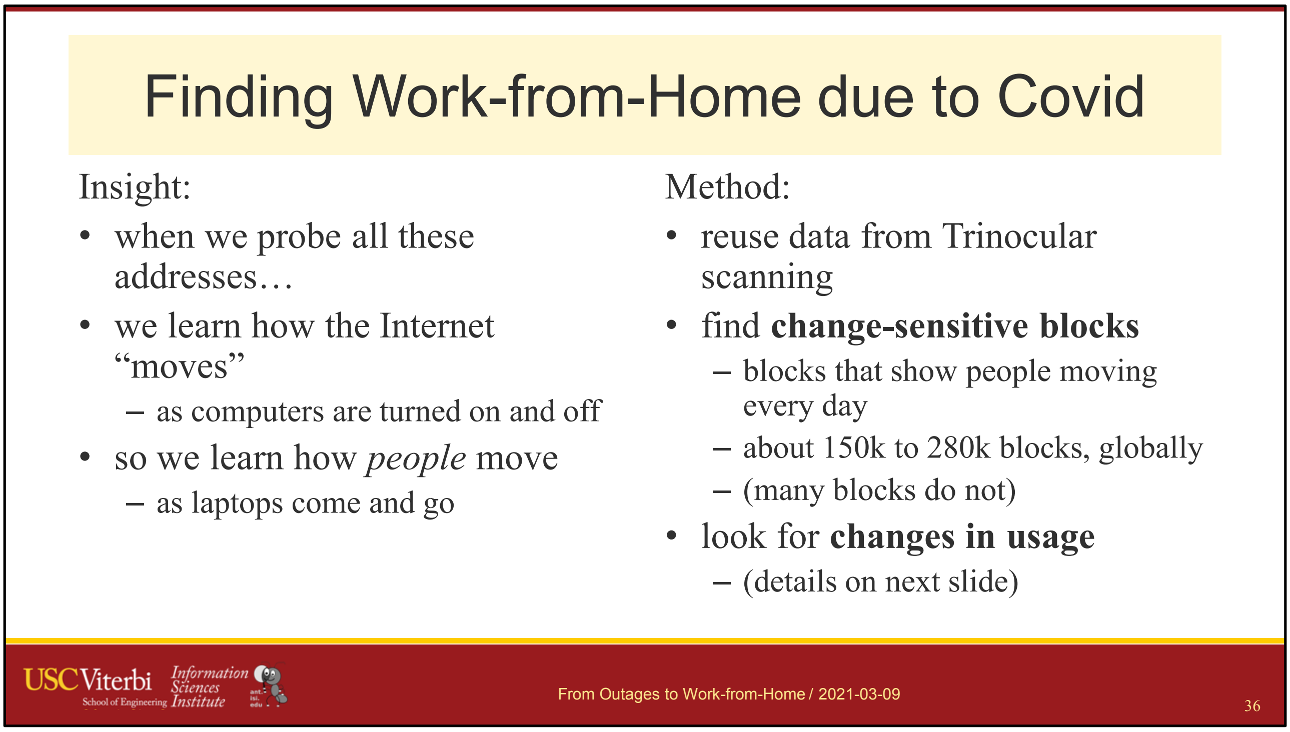
https://ant.isi.edu/diurnal/ani/
(play)

2013-04-26t09:11 (UTC)
r181
change from mean
+20%
+16%
+12%
+8%
+4%
0%
-4%
-8%
-12%
-16%
-20%
land (no networks)
ocean (no networks)
http://www.isi.edu/ant/diurnal/
internet_outage_adaptive_a12_diurnal-20130424
Copyright (C) 2014 by University of Southern California

fewer: pre-dawn in South America

more: afternoon in India

Details in "When the Internet Sleeps: Correlating Diurnal Newtorks with External Factors", by Quan, Heidemann, Pradkin in ACM IMC 2014. https://doi.org/10.1145/2663716.2663721



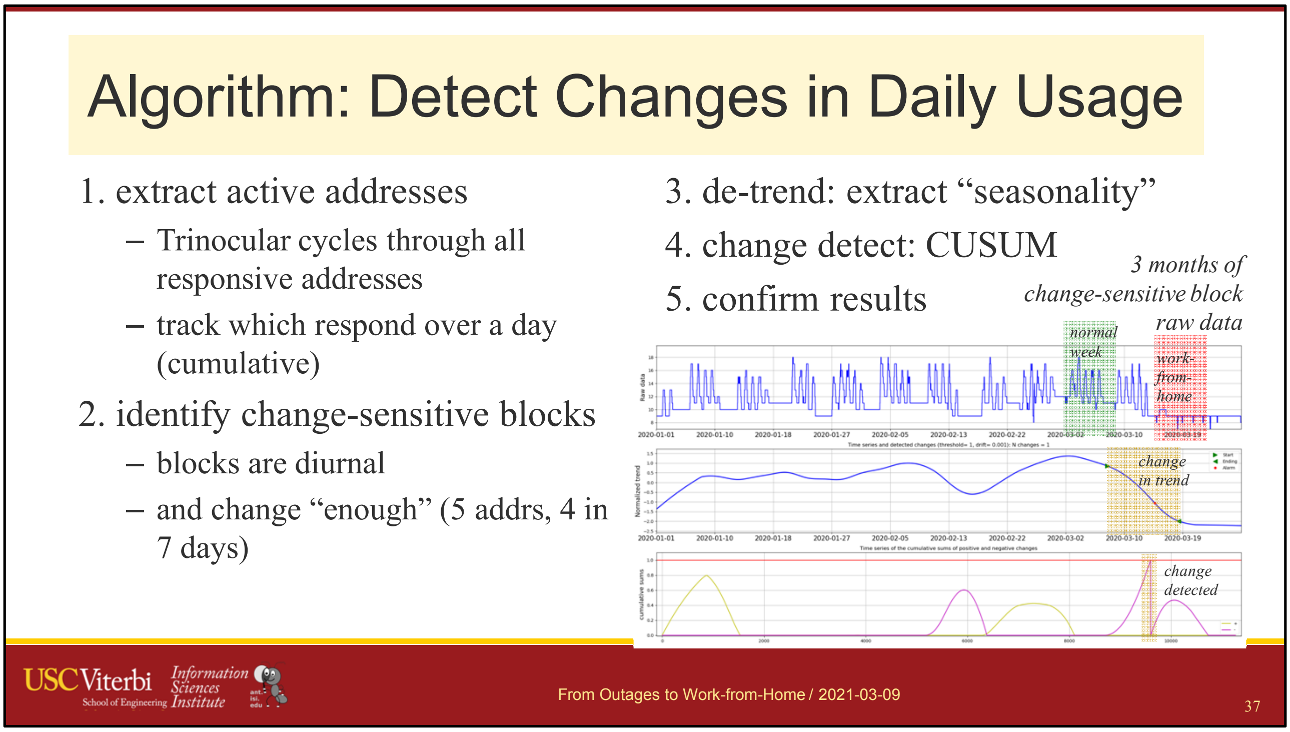
Finding Work-from-Home due to Covid

Insight:

- when we probe all these addresses…
- we learn how the Internet "moves"
  - as computers are turned on and off
- so we learn how *people* move
  - as laptops come and go

Method:

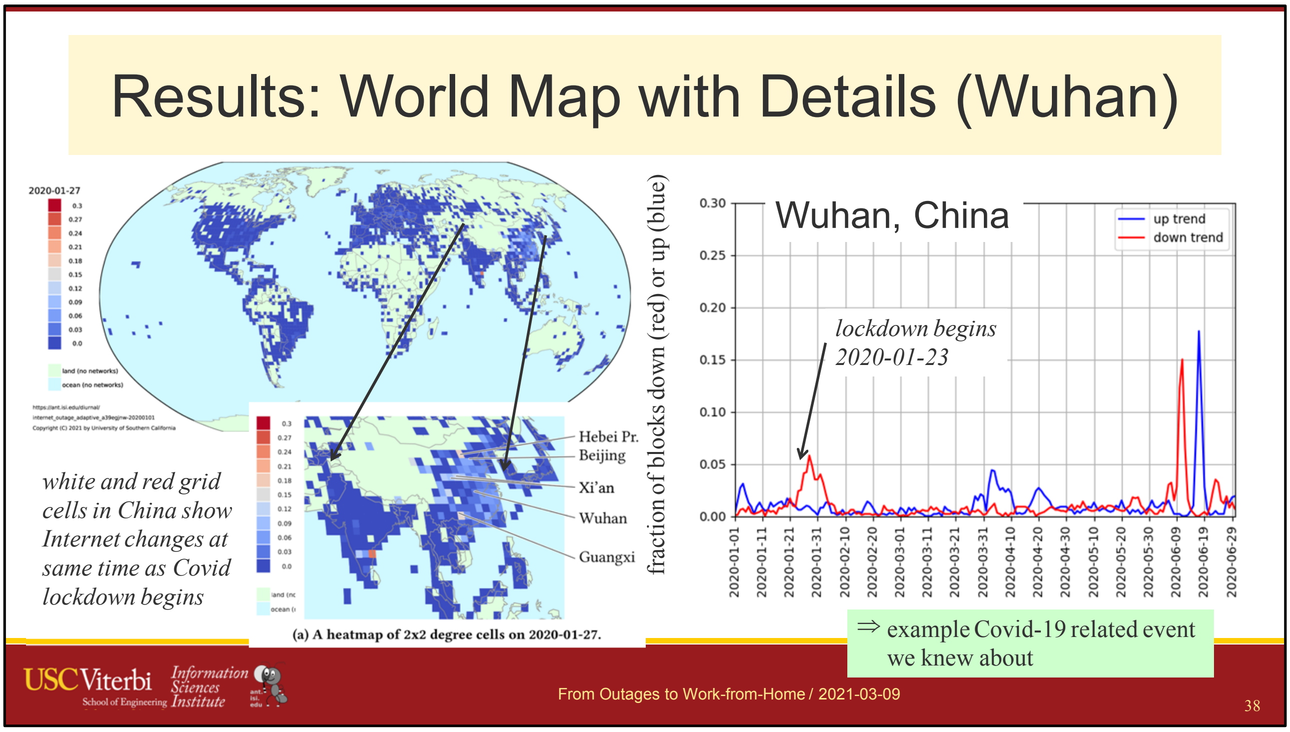
- reuse data from Trinocular scanning
- find **change-sensitive blocks**
  - blocks that show people moving every day
  - about 150k to 280k blocks, globally
  - (many blocks do not)
- look for **changes in usage**
  - (details on next slide)



Algorithm: Detect Changes in Daily Usage

1. extract active addresses
   - Trinocular cycles through all responsive addresses
   - track which respond over a day (cumulative)
2. identify change-sensitive blocks
   - blocks are diurnal
   - and change “enough” (5 addrs, 4 in 7 days)
3. de-trend: extract “seasonality”
4. change detect: CUSUM
5. confirm results

3 months of change-sensitive block raw data



Results: World Map with Details (Wuhan)

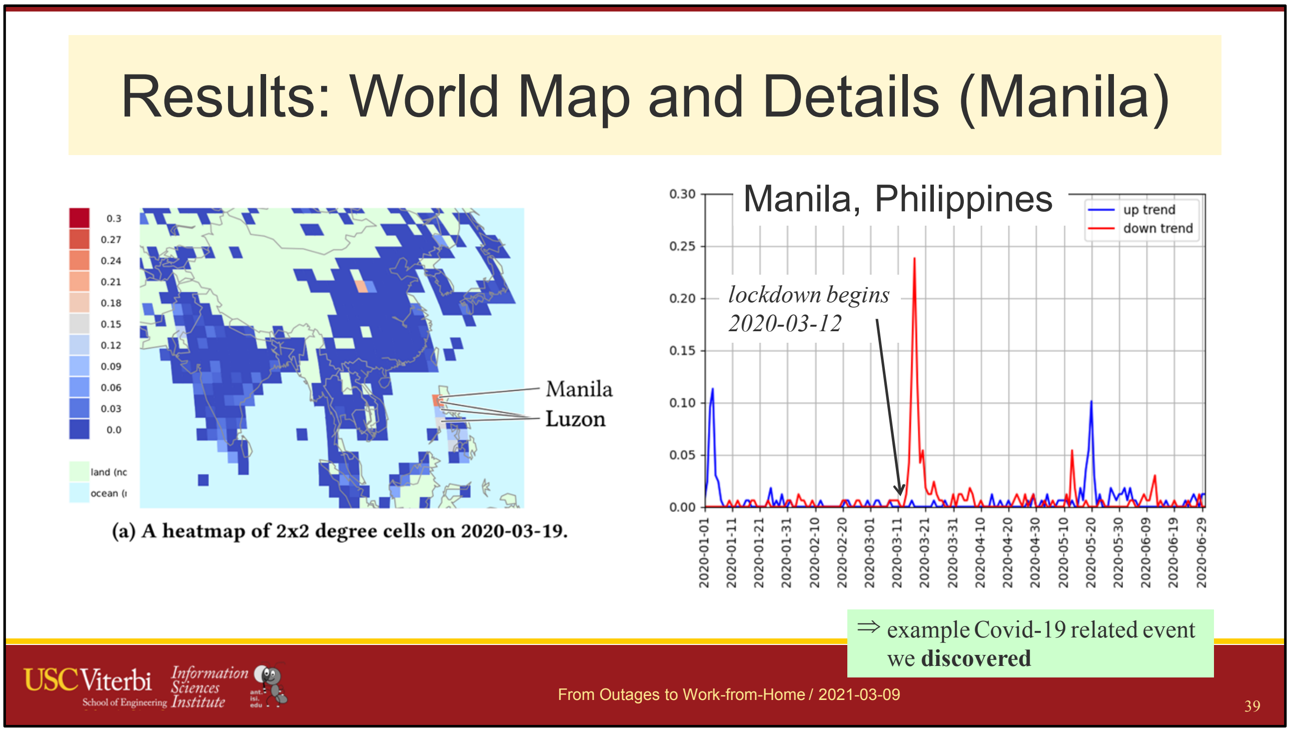
2020-01-27
0.3
0.27
0.24
0.21
0.18
0.15
0.12
0.09
0.06
0.03
0.0
land (no networks)
ocean (no networks)
https://ant.isi.edu/diurnal/
internet_outage_adaptive_a39egjnw-20200101
Copyright (C) 2021 by University of Southern California
Hebei Pr.
Beijing
Xi’an
Wuhan
Guangxi

white and red grid cells in China show Internet changes at same time as Covid lockdown begins

(a) A heatmap of 2x2 degree cells on 2020-01-27.

lockdown begins 2020-01-23

⇒ example Covid-19 related event we knew about



Results: World Map and Details (Manila)

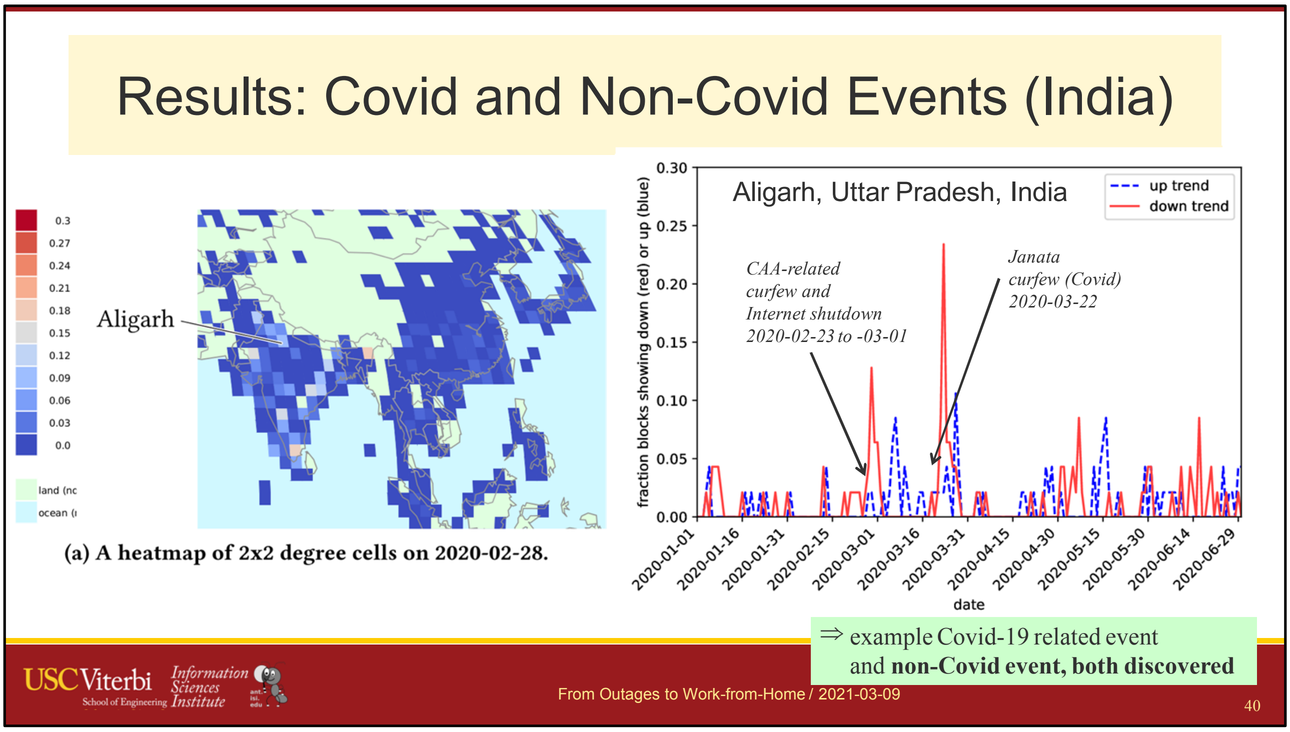
(a) A heatmap of 2x2 degree cells on 2020-03-19.

⇒ example Covid-19 related event we **discovered**

USC Viterbi School of Engineering Information Sciences Institute

From Outages to Work-from-Home / 2021-03-09

39



Results: Covid and Non-Covid Events (India)

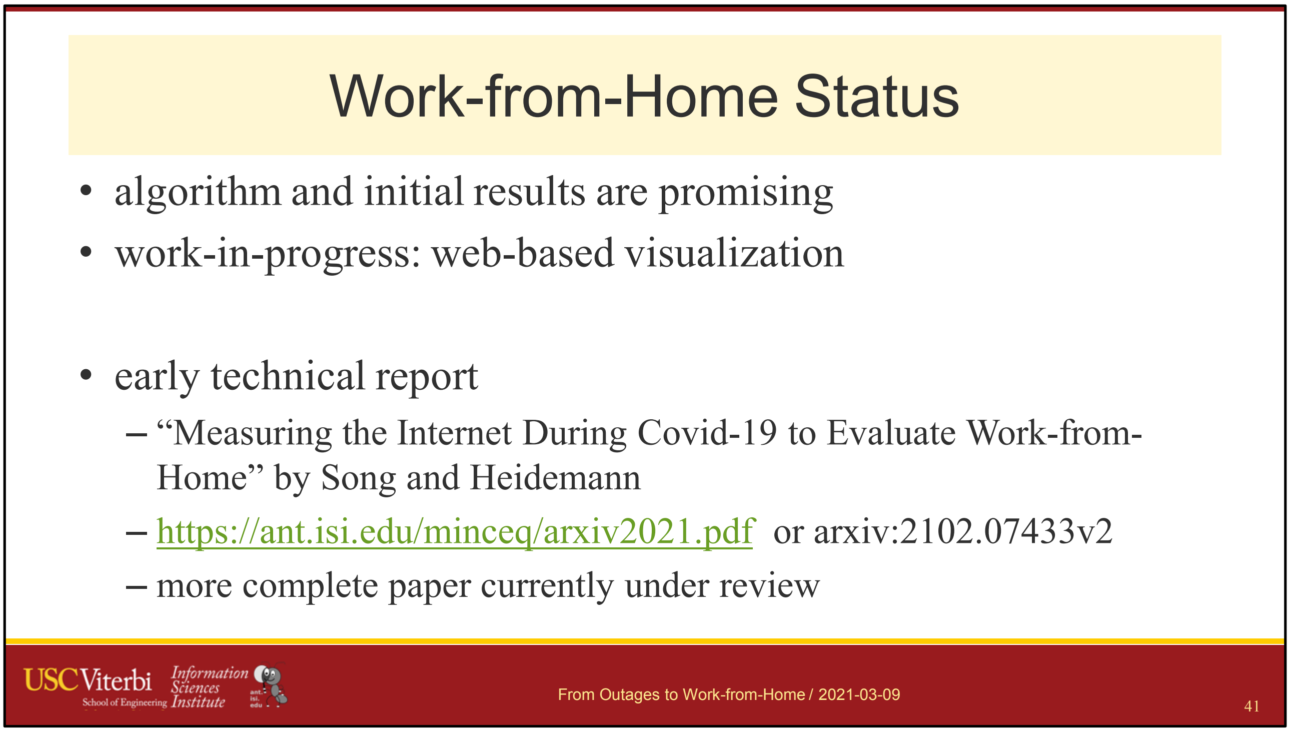
(a) A heatmap of 2x2 degree cells on 2020-02-28.

⇒ example Covid-19 related event and **non-Covid event, both discovered**

USC Viterbi School of Engineering Information Sciences Institute

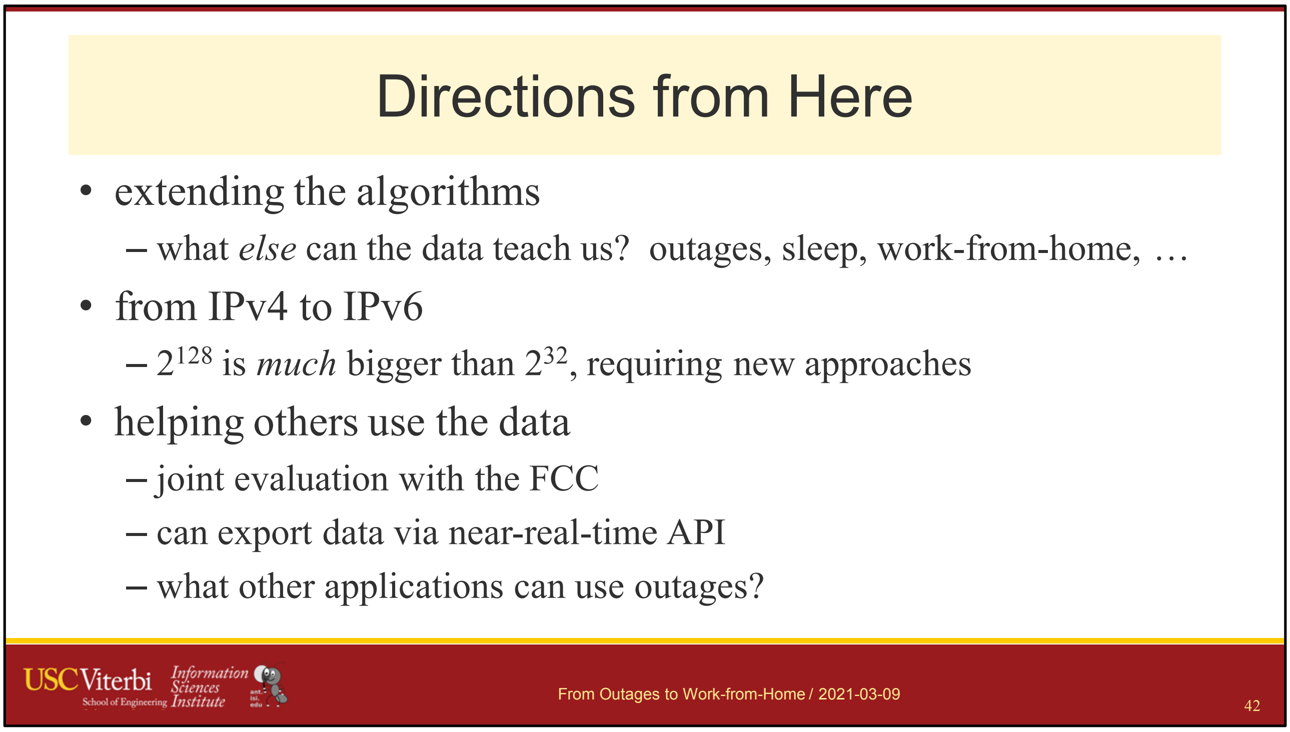
From Outages to Work-from-Home / 2021-03-09

40



Work-from-Home Status

- algorithm and initial results are promising
- work-in-progress: web-based visualization
- early technical report
  - “Measuring the Internet During Covid-19 to Evaluate Work-from-Home” by Song and Heidemann
  - https://ant.isi.edu/minceq/arxiv2021.pdf or arxiv:2102.07433v2
  - more complete paper currently under review

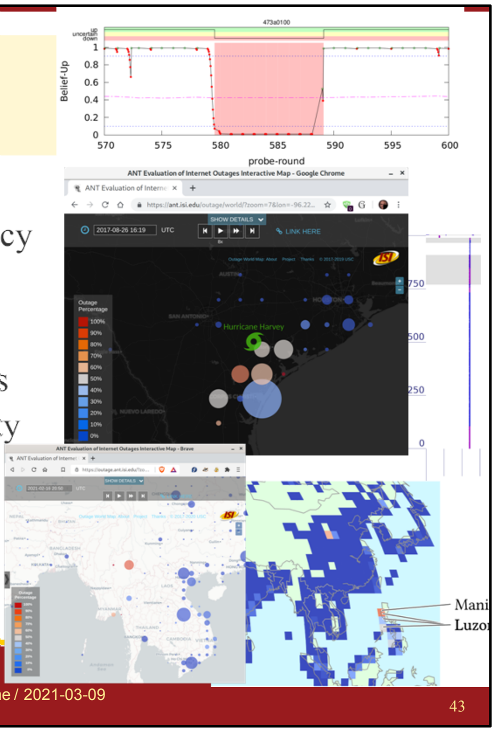


Directions from Here

- extending the algorithms
  - what *else* can the data teach us? outages, sleep, work-from-home, …
- from IPv4 to IPv6
  - 2^{128} is *much* bigger than 2^{32}, requiring new approaches
- helping others use the data
  - joint evaluation with the FCC
  - can export data via near-real-time API
  - what other applications can use outages?

Conclusions

- we *can* measure Internet outages
  - precisely: for millions of nets; ~11-minute accuracy
  - in near-real time
- outages have many applications:
  - short-term: helping first responders, ISPs, citizens
  - long-term: understanding and improving reliability
- looking for partners and data consumers
- more info? papers and data https://ant.isi.edu/
  - maps: https://outage.ant.isi.edu/


473a0100
up
uncertain
down
Belief-Up
1
0.8
0.6
0.4
0.2
0
570
575
580
585
590
595
600
probe-round
ANT Evaluation of Internet Outages Interactive Map - Google Chrome
2017-08-26 16:19
UTC
SHOW DETAILS
LINK HERE
Outage Percentage
Hurricane Harvey
ANT Evaluation of Internet Outages Interactive Map - Brave
Mani
Luzo