IODA NP: Internet Outage Detection and Analysis

Alberto Dainotti | CAIDA, UC San Diego

Nov 13, 2019

Homeland Security
Science and Technology

Team Profile

- PI: Dr. Alberto Dainotti – Research Scientist
- Team: *Alistair King, Rama Padmanabhan, Dan Andersen, Paul Hick, Alex Ma, Marina Fomenkov*

- CAIDA - Center for Applied Internet Data Analysis
University of California, San Diego

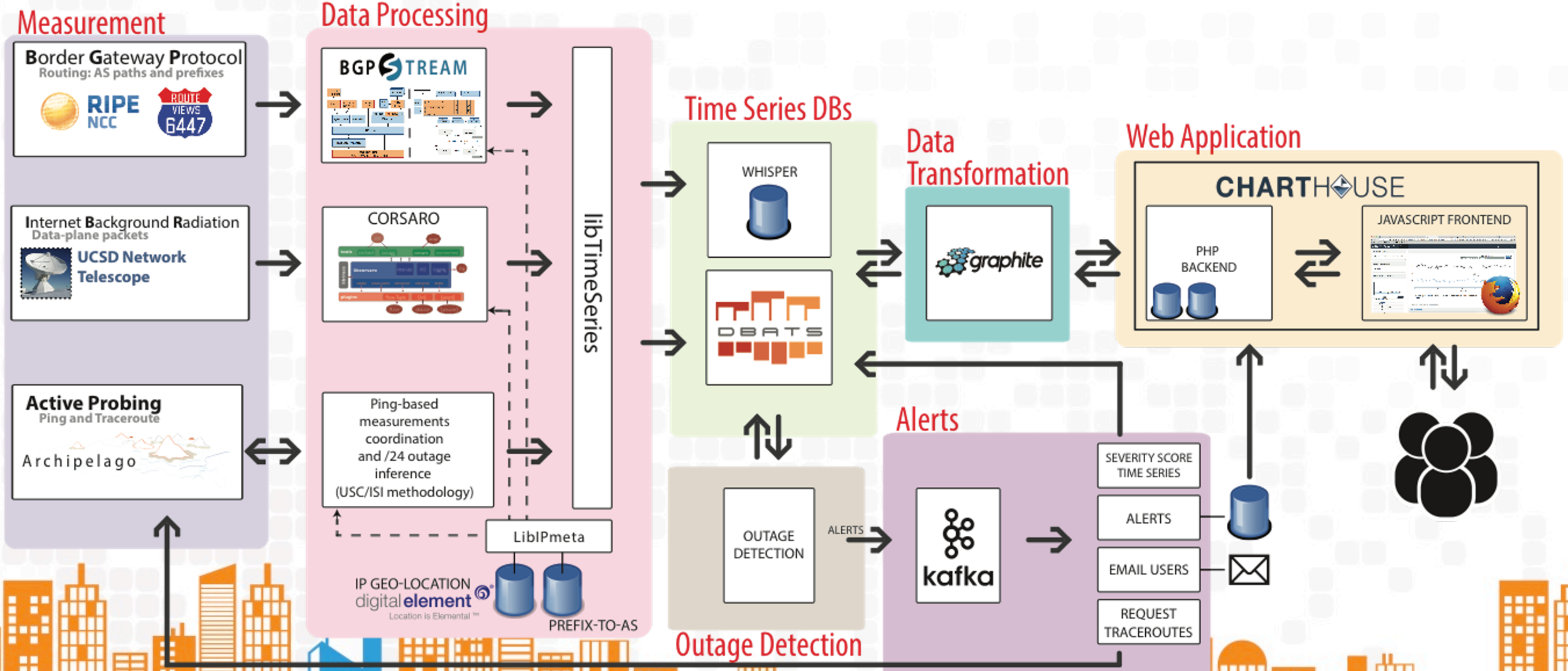
Customer Need

- Goal: Timely Detection and Analysis of Internet Connectivity Outages
  - Focus on macroscopic events, affecting the network edge
  - *E.g., a connectivity black-out significantly affecting customers of a large network operator or a large geographic area*
- Context: Cyber attacks, physical attacks, natural disasters, bugs and misconfiguration, government orders, …
- Application: Public Safety, Situational Awareness, Disaster Recovery, Insurance, Internet Reliability & Performance

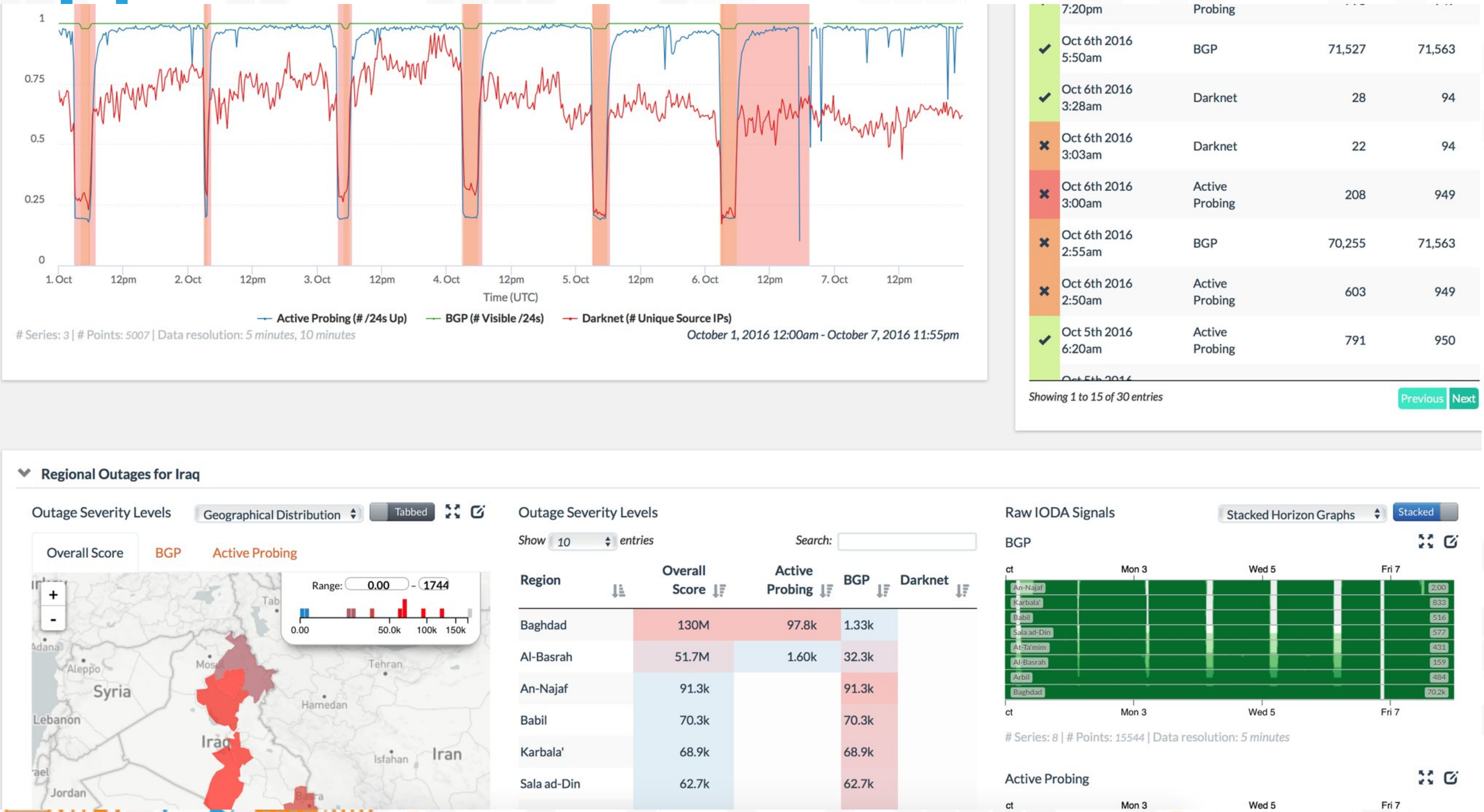
Approach Overview

- IODA: Internet Outage Detection & Analysis
  - Combine *active* and *passive* measurements both at the *data plane* and *control plane*
  - Data aggregation and event detection per Autonomous System (AS) and Geographic Area
  - Interactive Visualization
- IODA-NP: *Next Phase*
  - Methodological improvements and evaluation based on ground-truth and cross-validation
  - Reporting events
  - API framework and documentation

Approach - Architecture



Approach - Visualization

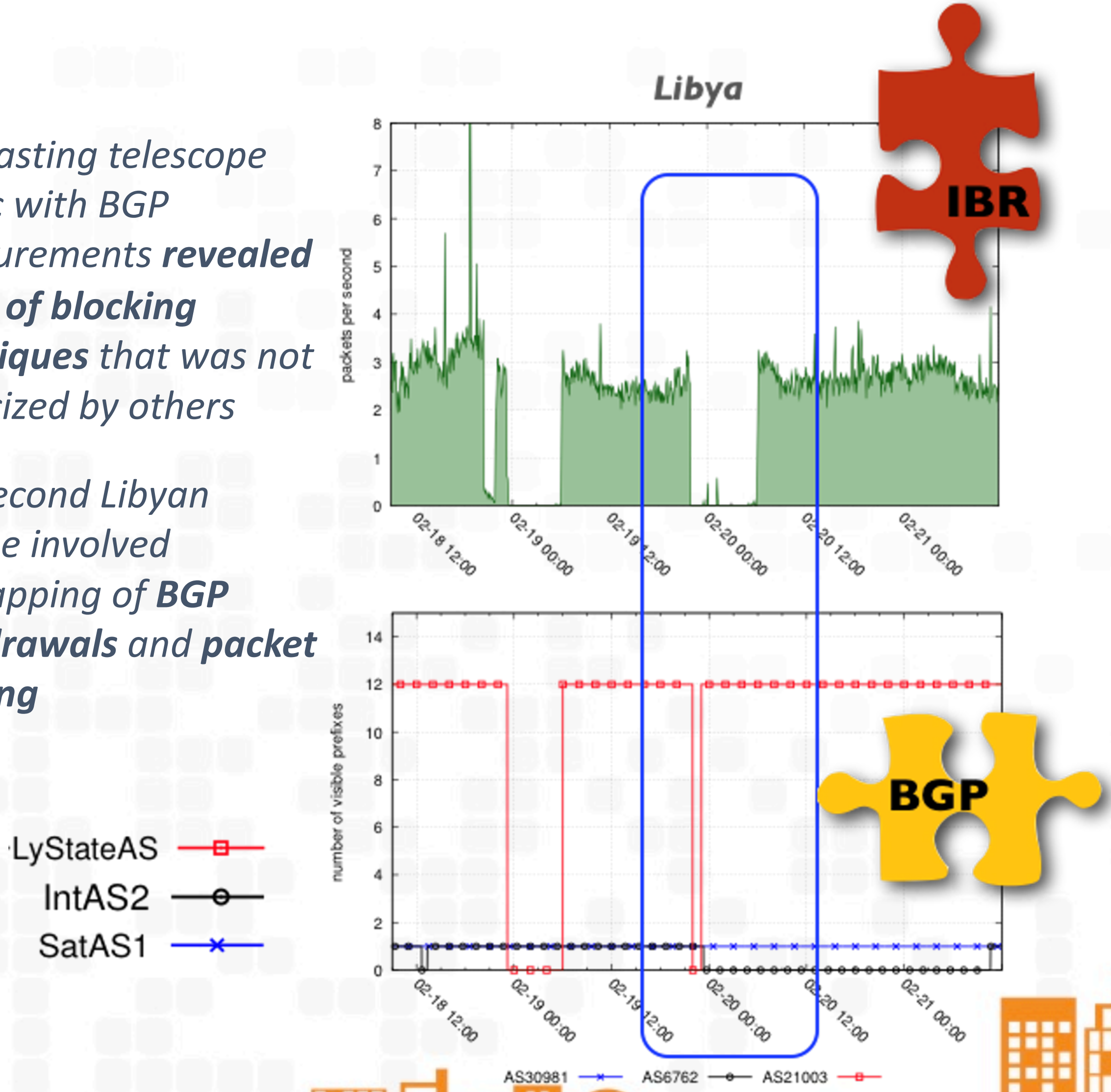


Benefits

- **Benefits**
  - Near-realtime alerts
  - Multi-source
  - Visualization
  - API
- **Risks / Challenges**
  - Complex infrastructure
  - Improve resolution
  - Validation

*Contrasting telescope traffic with BGP measurements **revealed a mix of blocking techniques** that was not publicized by others*

*The second Libyan outage involved overlapping of **BGP withdrawals** and **packet filtering***



Competition/Alternatives

- Oracle's Internet Intelligence Map
  - Focus on country-level
  - Limited interaction/viz functionalities
- ISI / John Heidemann's work
  - IODA uses ISI Trinocular for one data source
  - IODA focuses on per-AS / geographic aggregations
- Akamai
  - State of the Internet reports and tweets
- Bgpmon.com
  - BGP only

Accomplishments (1/2)

- Methodological **improvements** for existing inference methods
  - BGP: Improved our approach for prefix geolocation (TMA '19)
  - IBR: Developed a new approach for outage (change point) detection (TMA'19)

- **New** methodologies and investigations
  - Detecting outages at finer granularity
  - Quantifying the effect of weather upon residential Internet (SIGCOMM '19)
  - Investigation of cross/partial /24 outages (PAM '19)
  - Active Probing: development of a complementary technique (WIP: Zeusping)

Accomplishments (2/2)

- Architecture/DevOps improvements
  - Purchase and deployment of new hardware to support IODA
  - Streamlined active probing management
  - Inferring outages from the IBR data source in near-realtime
  - Automated deployment & configuration management
  - InfluxDB as a potential replacement for DBATS
  - Redesigned approach to handle level-shifts in time series
  - New API for IODA's Alerts
- Towards evaluation against ground truth

Methodological improvements: BGP [TMA'19]

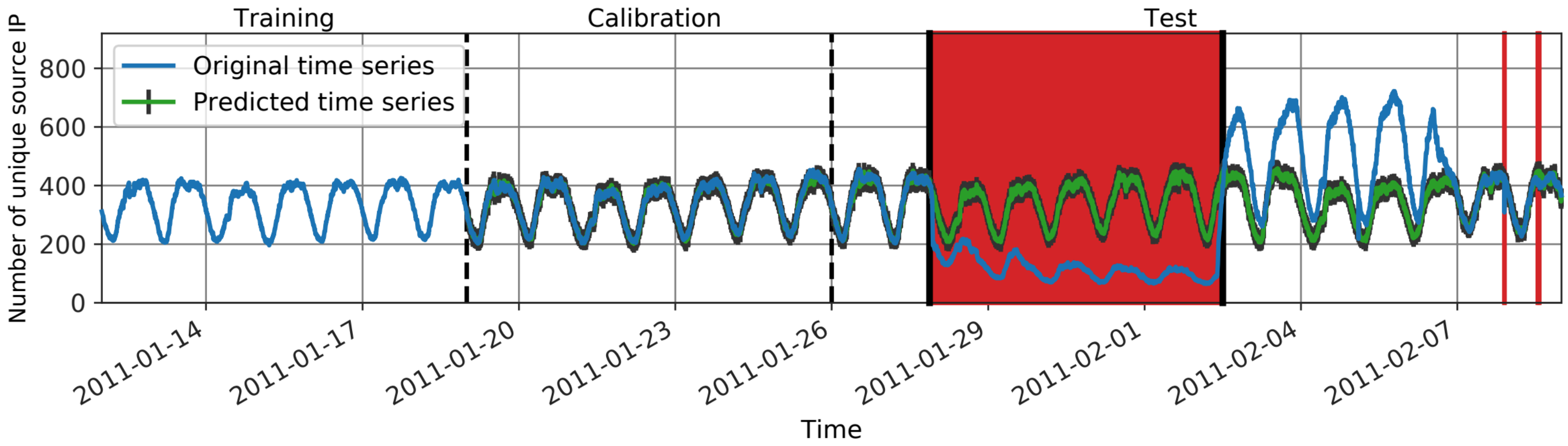
- We improved our approach for geolocating BGP prefixes
  - Some BGP prefixes geolocate to multiple locations: 2% of prefixes map to multiple countries and 27% to multiple cities
  - Our prior approach overestimated outages when such prefixes were withdrawn
  - We improved our approach to better quantify the extent of the outage in different locations when such a prefix was withdrawn
- We identified a new problem: prefix geolocation ambiguity
  - BGP prefixes that *contain* other prefixes can have ambiguous geolocation
  - When such prefixes are withdrawn, it is challenging to identify which locations were affected
  - We studied and quantified the extent of ambiguity in the global routing table
  - We identified preliminary features that could help resolve the ambiguity (e.g.: different origin ASes)
- Outcome: Increased the accuracy of outages detected using BGP
- TMA '19 paper: https://www.caida.org/publications/papers/2019/geo-locating_bgp_prefixes/geo-locating_bgp_prefixes.pdf

Methodological improvements: IBR [TMA'19]

- IODA currently detects outages for the IBR data source using conservative thresholds
  - The IBR signal exhibits large fluctuations
  - We missed detecting many outages

- We developed a novel technique for outage detection (collab. IIJ & Univ. Strasbourg)
  - Using historical data from the network telescope, the technique employs Seasonal ARIMA (S-ARIMA) to model IBR traffic
  - This model is then used to predict future IBR traffic: when the observed traffic is lower than the predicted traffic, an outage is detected

S-ARIMA-based outage detection



Methodological improvements: IBR

- **Prototype**:
  - Developed a prototype and evaluated the S-ARIMA-based approach against our current approach
  - The new approach outperforms the current one, detecting two orders of magnitude more outages
  - We published this approach and initial results in TMA '19
  - TMA '19 paper: https://www.caida.org/publications/papers/2019/chocolatine/chocolatine.pdf
- **Full development, integration, deployment:**
  - Completed the development and initial deployment of the S-ARIMA-based approach for outage detection with IBR
- **Testing**:
  - We are currently monitoring its performance in terms of stability and data validity
  - Once the testing phase is completed, we plan to switch IODA's production alerting system to this new algorithm

Detecting outages at finer geographic granularity

- IODA currently detects outages at the country-level, region-level, and AS-level
- But some outages may only be visible at finer geographic granularity
  - E.g.: Outages due to localized weather events such as tornados
- We performed initial investigations into detecting outages at fine geographic granularities in the U.S.
  - Specifically, we investigated whether tornado-related outages are visible in IODA's **county-level time series signals**

Investigating tornado-related outages at the county-level

NATIONAL

'Tornado Outbreak' Devastates Ohio Communities With Winds Up To 140 MPH

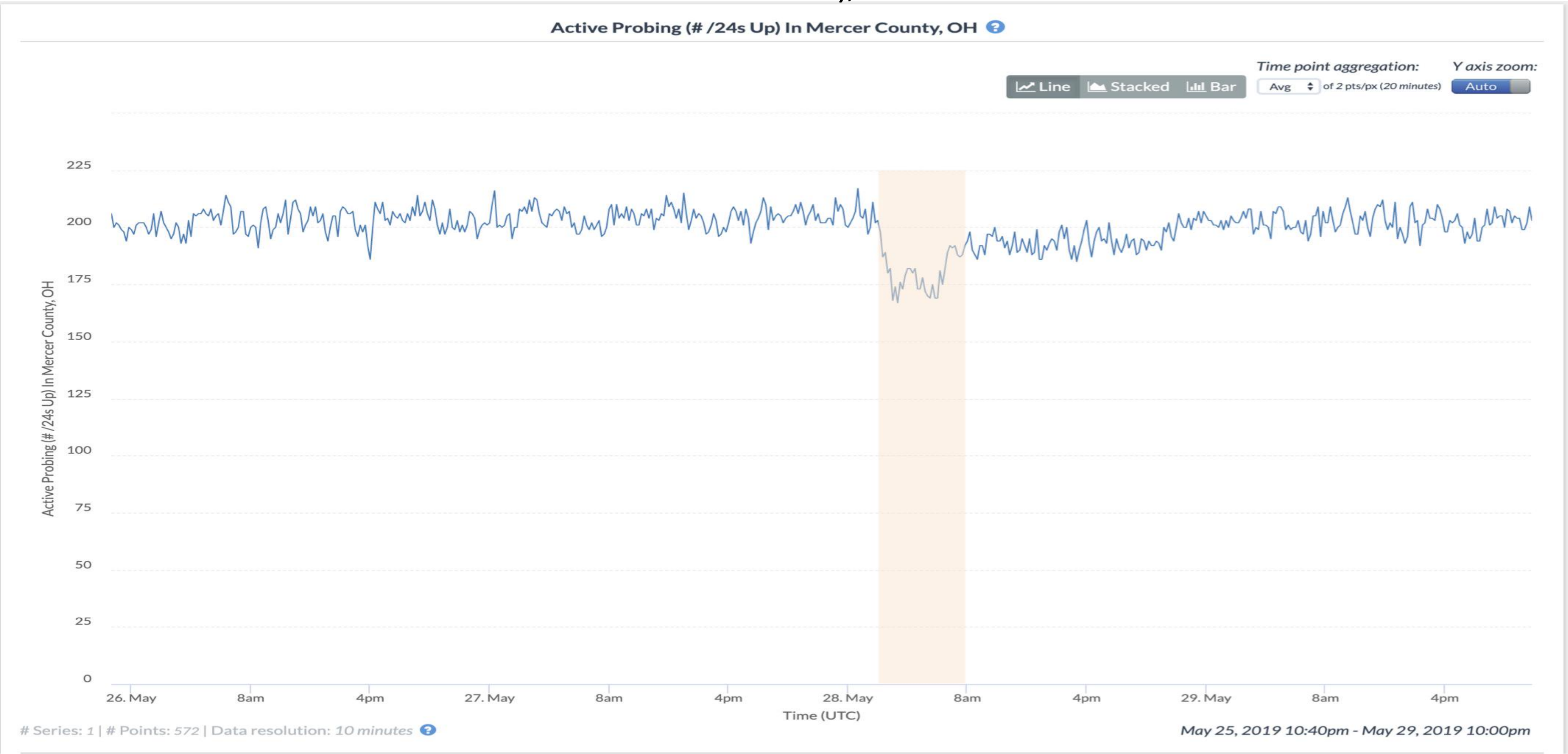
May 28, 2019 · 2:43 AM ET

VANESSA ROMO   BILL CHAPPELL

March 28 2019

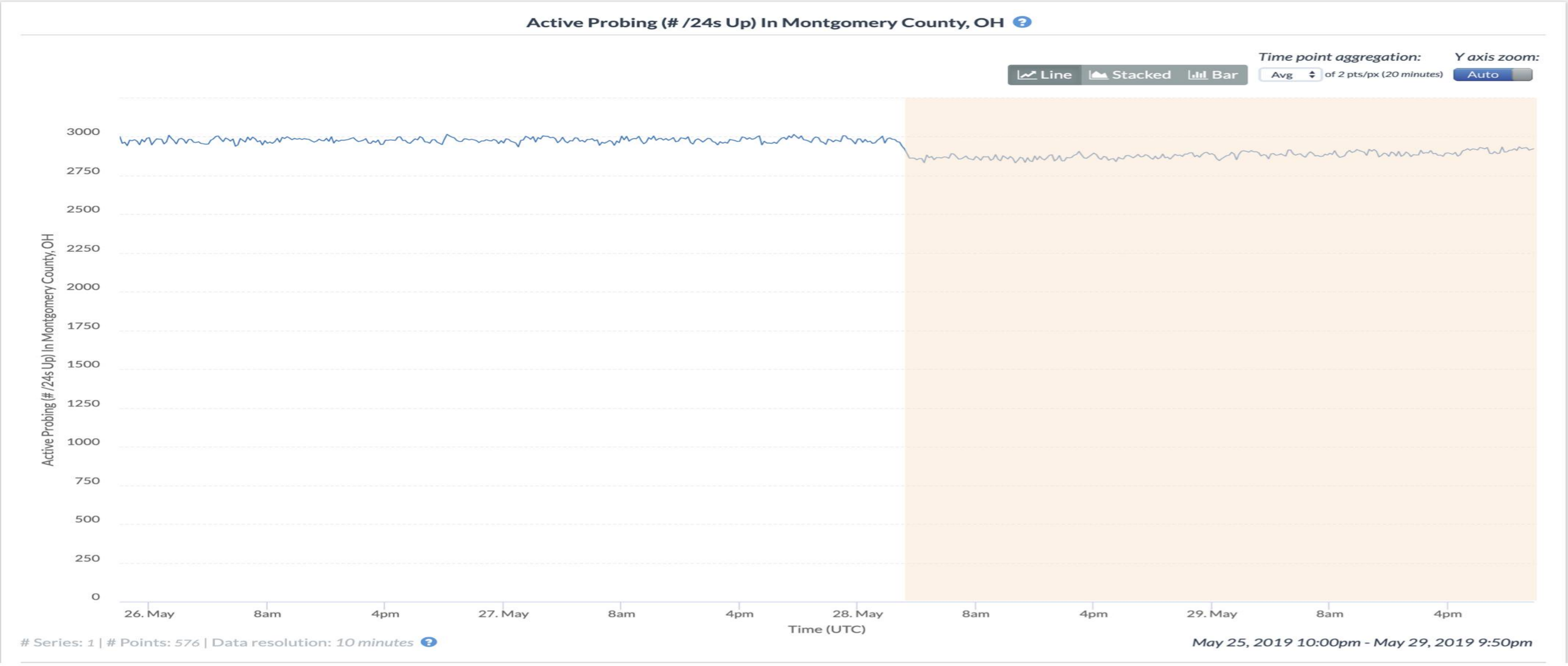
Investigating tornado-related outages at the county-level

Mercer County, OH



Investigating tornado-related outages at the county-level

Montgomery County, OH



Investigating tornado-related outages at the county-level

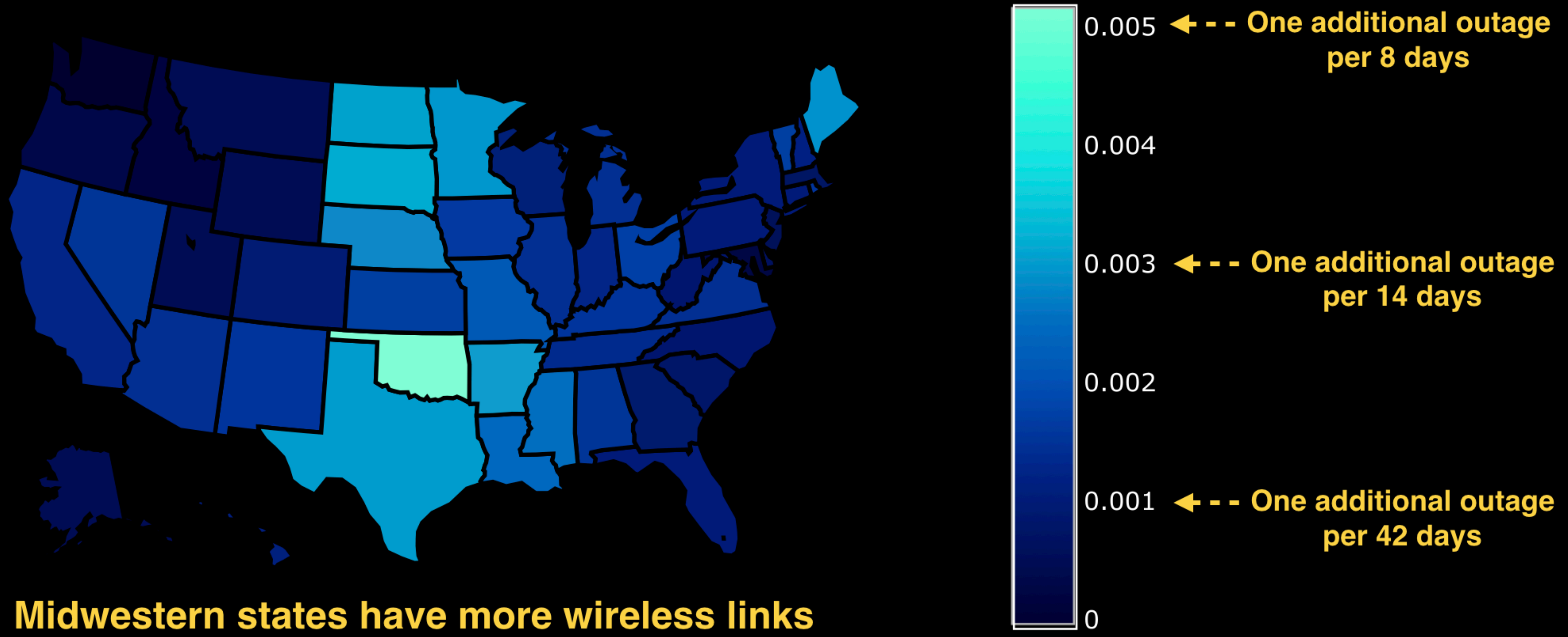
- We were able to correlate a few reports of tornadoes with drops in IODA's time series signals
  - Typically, these drops occurred for the active probing inference method

- However, most tornado reports did not correlate with significant drops in IODA's time series signals

Quantifying the effect of weather using a complementary dataset [SIGCOMM'19]

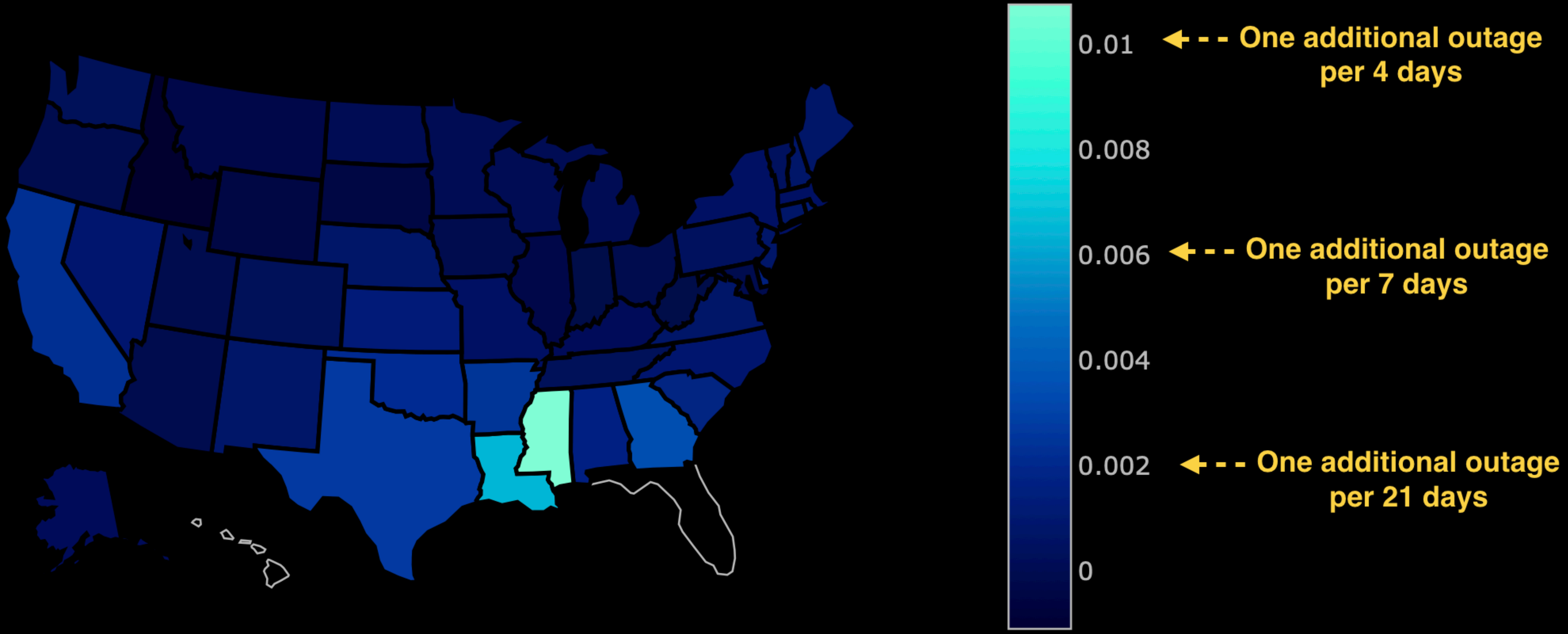
- We used 8 years of data collected by UMD's Thunderping project to analyze how weather affected residential Internet in the U.S.
  - Thunderping uses forecasted weather alerts from the NWS to identify counties that could experience severe weather
  - Samples individual addresses from multiple providers in these counties
  - Pings them from multiple Planetlab vantage points
- Diverse weather conditions are correlated with increased outage probability
- The effects of weather conditions vary depending upon media-type, geography, and intensity
- SIGCOMM '19 paper: https://www.caida.org/publications/papers/2019/residential_links_under_weather/residential_links_under_weather.pdf

Midwestern states are more vulnerable to rain

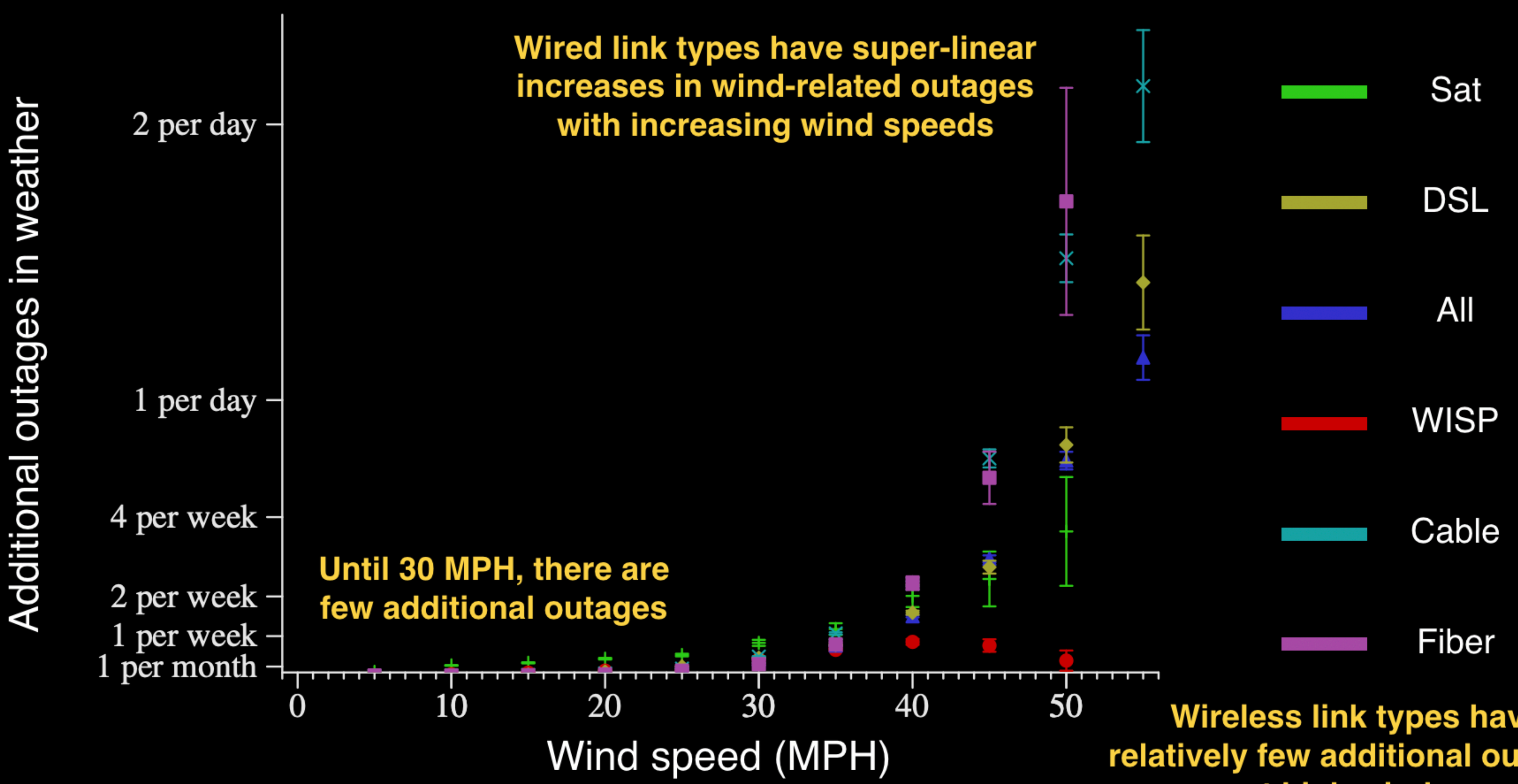


Midwestern states have more wireless links

Southern states are more vulnerable to snow

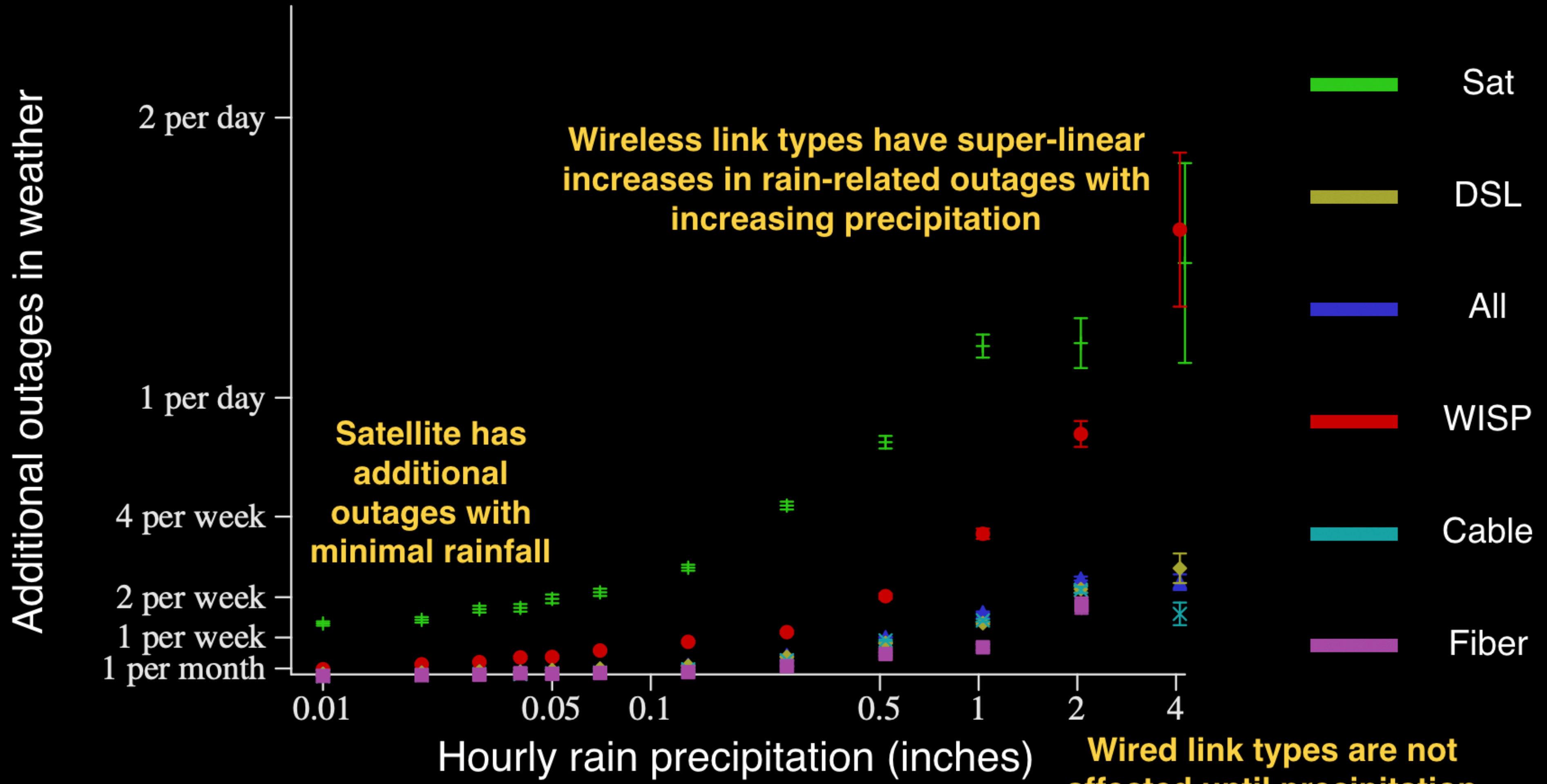


Wind affects wired links more than wireless links



Wireless link types have relatively few additional outages even at high wind speeds

Rain affects wireless links more than wired links



Wired link types are not affected until precipitation exceeds 0.5 inches

Why was the visibility of weather-related failures lower in IODA?

- **IBR** does not (yet) seem capable of catching such small events
- **BGP** does not either: the control-plane is most of the time unaffected
- **Active probing** detected only a few.
  - IODA's active probing data source uses the Trinocular methodology to detect outages affecting /24 blocks
    - The intuition is that neighboring addresses in /24 blocks would share fate
  - We investigated how often outages affect entire /24 blocks using data from the complementary Thunderping methodology

Identify correlated outages in Thunderping dataset [PAM'19]

- Thunderping pings (sampled) individual addresses in small geographic areas like a county
- Dropout event: when an address that was responding to pings in the previous round stops responding to pings from all vantage points

- Individual vs Simultaneous dropouts:
  - If an **individual** address has a dropout, it could be due to several causes
    - E.g., users turning off their home router
  - **Simultaneous** dropouts of multiple related addresses could be due to a common underlying cause
    - E.g., power or network failures

- We developed a statistical technique using the Binomial test to identify **correlated dropouts** that are highly unlikely to have occurred independently
  - Correlated dropouts are likely due to outage events

Investigate cross/partial /24 outages [PAM'19]

- For correlated outages in Thunderping, we checked how often ***all*** addresses in the /24 were affected
- 69% of /24s that had at least one address affected by an outage event had at least one other address that was not affected
- E.g.: in 1 event, 811 addresses from 42 /24 blocks had a correlated failure but 40 of the /24 blocks had at least one address that continued to respond to pings

- Outcome: Even large outage events may affect (multiple) /24 blocks partially
- PAM '19 paper: https://www.caida.org/publications/papers/2019/how_find_correlated_internet/how_find_correlated_internet.pdf

A new (complementary) active probing technique: Zeusping

- The Thunderping dataset allowed us to perform preliminary investigations of cross/partial /24 outages but has a fundamental limitation
  - It only measures sampled addresses and only during forecasted weather events

- Our goal with this technique is to characterize outages
  - IP address dimension: How are addresses affected by an outage event related in the IP address space?
  - Geographic dimension: How are the addresses affected by an outage event related by geography?
  - Time dimension: How long do outages last?

Complementary active probing technique: Zeusping

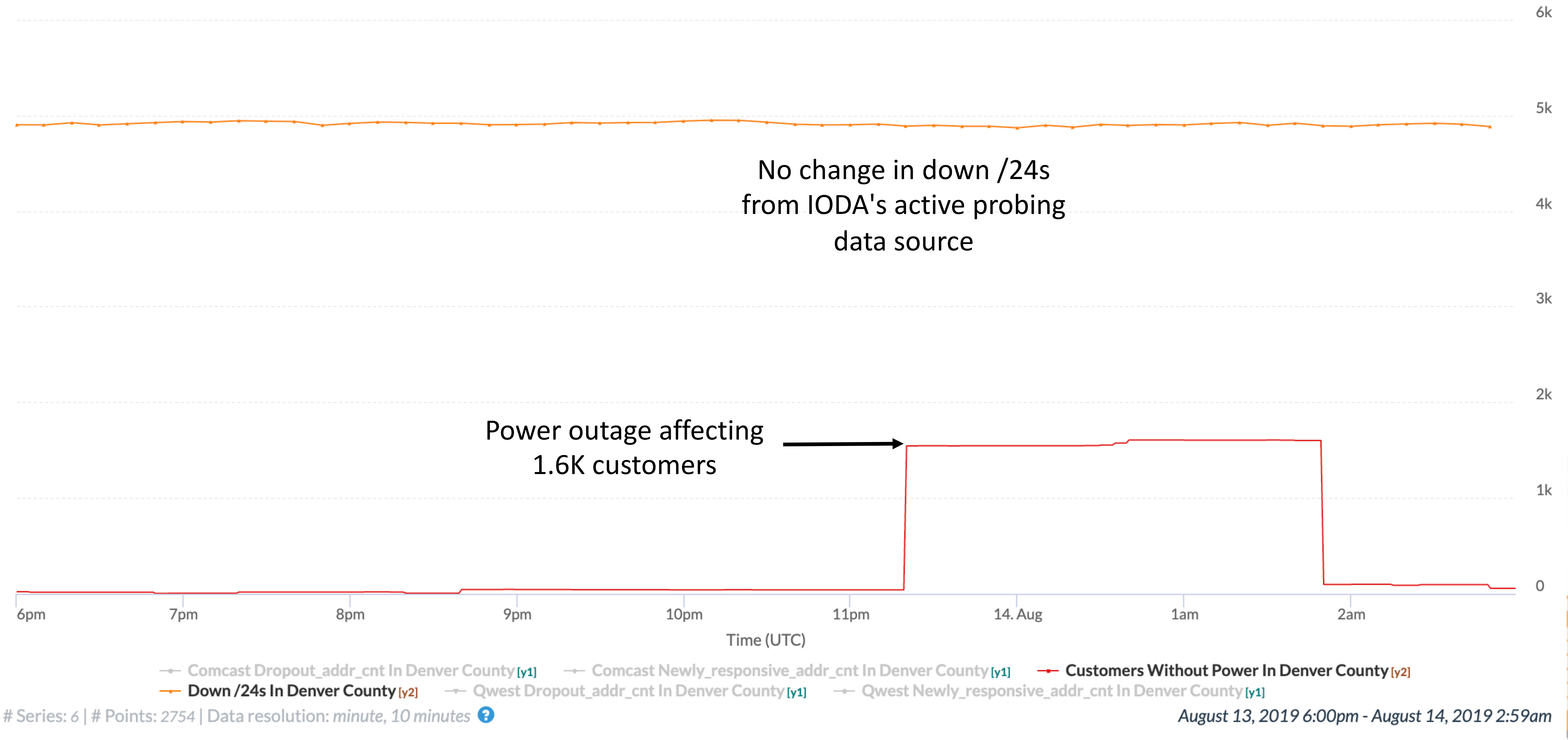
- Approach
  - Probe all addresses in a geographic region
  - Each address receives a ping every 10 minutes from three AWS vantage points
  - Identify a dropout when an address that is responding to pings in the previous round stops responding to pings from all vantage points in the next round
  - Use simultaneous dropouts to identify a correlated outage event

Zeusping: Preliminary results

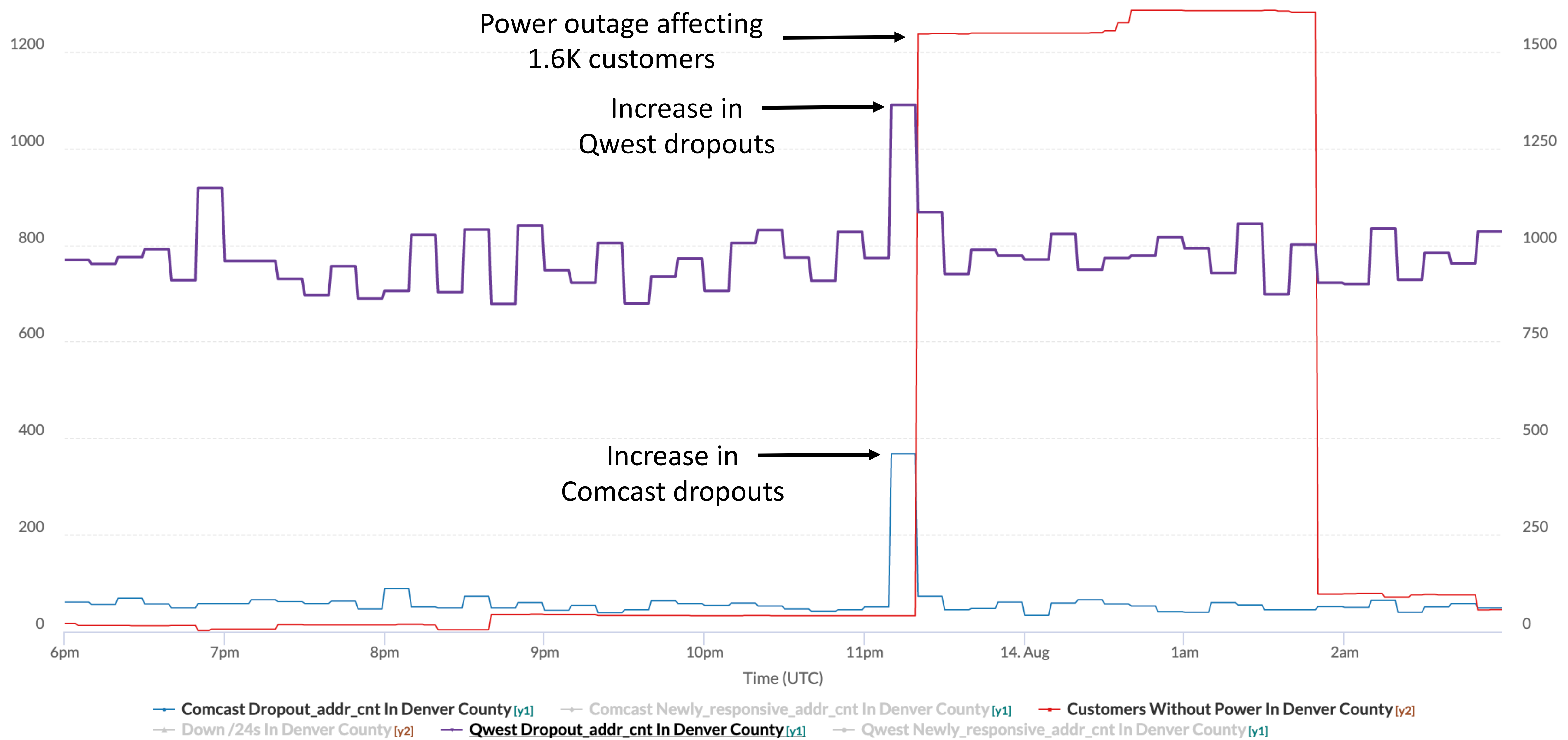
- We chose Colorado as the starting point for Zeusping, since we have a source of ground truth power outages in the state
  - Xcel Energy, the largest power provider (estimated to be ~80%) in Colorado, displays live power outages on their website
  - https://www.outagemap-xcelenergy.com/outagemap/?state=CO
  - We have been scraping this data since March 2019, loading it into IODA

- Probed 4M addresses in Colorado since mid-August
  - 1.7M Charter addresses
  - 1.2M Comcast addresses
  - 1.1M Centurylink addresses

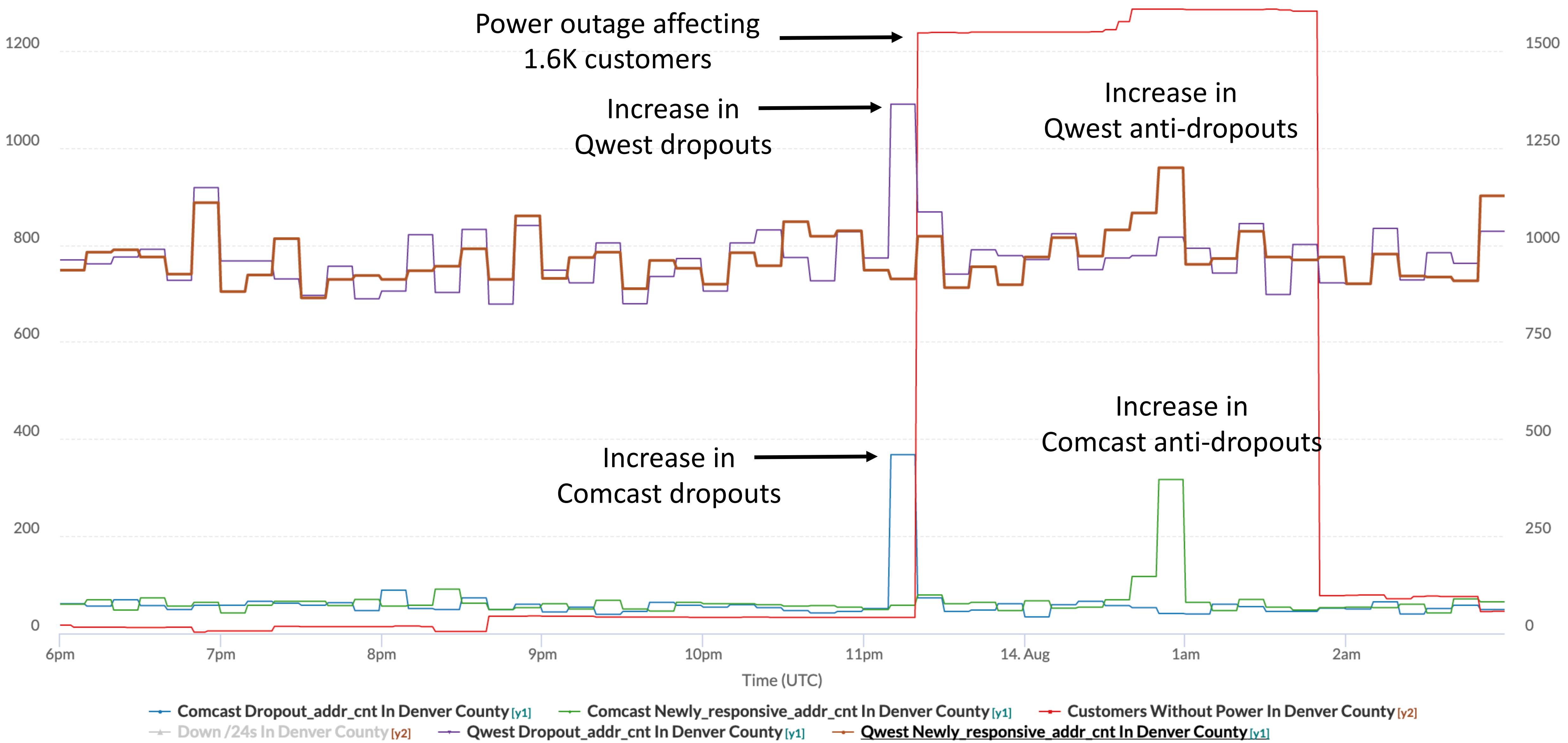
Power outage ground truth vs. IODA's active probing



Power outage ground truth vs. Zeusping



Power outage ground truth vs. Zeusping



Purchase and deployment of new hardware to support IODA

- Purchased and deployed an SSD cluster node
  - Part of time series collection and storage cluster
  - Configured to store and serve IODA time series data

- Purchased and deployed an object storage server
  - Added to CAIDA's OpenStack Swift object storage cluster
  - This cluster is used for storing raw IODA measurement data
  - Total capacity of Swift cluster is now 1.65 PB

Streamlined active probing management

- Configured two machines dedicated to active probing measurements
- Streamlined the process of generating and deploying new probe-lists
  - Probe-lists contain the addresses that will be probed
  - Automated the generation and distribution of new probe-lists to probers
- Outcome: Increased the reliability of our infrastructure for active probing

Inferring outages from the IBR data source in near-realtime (1/2)

- Previously, IODA's outage detection with IBR had a theoretical minimum detection latency of 1 hour
  - Data was processed in 1-hour batches, resulting in this theoretical minimum
  - In practice the latency was usually ~2hrs, sometimes days (in the aftermath of system failures)

Inferring outages from the IBR data source in near-realtime (2/2)

- Developed state-of-the-art realtime data-collection, distribution, and analysis systems
  - Deployed new 10Gbps-capable packet capture and distribution software
  - Rearchitected Corsaro and time series plugin to be highly parallel and efficient
  - Tested and deployed this system into our production environment
- Outcome: Reduction of outage detection latency from multiple hours to <2 minutes

Automated deployment & configuration management

- Goal: Automatically deploy appropriate configuration files to different Virtual Machines (VMs) running IODA's front-end web application
  - E.g.: when a Github change is pushed, the VMs automatically deploy the latest code
- Implemented a new (Puppet-based) configuration management environment
  - Completed a prototype environment for the IODA HTTP API service
  - Ported all components of the IODA web application infrastructure
- Outcome: Improves the reliability (e.g., reduced downtime) and maintainability of IODA

Redesigned our approach to handle level-shifts in time series

- Previously, level-shifts in the time series signals of IODA's data sources resulted in numerous incorrect outage alerts
- We performed a fundamental redesign of the architecture of IODA’s outage detection module: Watchtower
  - Investigated tradeoffs in how IODA deals with alerts
  - Rewrote the Watchtower component to make it more flexible and modular
- Outcomes:
  - Reduced outage detection latency to within the next data interval
  - Reduced "structural" false positives
  - Released as open source (*https://github.com/caida/watchtower-sentry*)

New API for IODA's alerts

- The Common Outage Data Format
  - We collaborated with PARIDINE performers (John Heidemann, Matthew Luckie, …) to produce a data format for exchanging information about Internet outages
  - https://www.isi.edu/publications/trpublic/pdfs/isi-tr-729.pdf

- We redesigned IODA's Alerts API
  - We examined various use-cases for IODA's alerts and considered different fields we should expose to users
  - We implemented the new API and ensured it emits alerts in the *Common Outage Data Format*
  - https://ioda.caida.org/ioda/api

Towards evaluation against ground truth

- We manually investigate known outage events in IODA's time series
  - Some known outage events are reported in the news media (E.g., recent censorship event in Iraq)
  - We also investigate reports about potential outage events from the Internet Freedom community
    - We have actively engaged with #KeepItOn and other Internet Freedom groups (Tibcert, internetshutdowns.in, Venezuela Inteligente)
  - We have also been scraping power outage data from U.S. power companies
    - We have been scraping 9 of the top 10 power companies' power outage datasets since June 2018
    - We have since added 21 more power companies

Current Status

- Prototype running
  - Web dashboard – https://ioda.caida.org/ioda/dashboard
  - Twitter - https://twitter.com/caida_ioda
  - Blogposts
  - Collaborations (more later)

Lessons Learned: what went well

- Outage detection with all data sources now occurs in near real-time
- Performed important methodological and development improvements
- Increased uptake of IODA, several collaborations

Lessons learned: Remaining challenges

- IODA currently lacks visibility into networks that use Carrier-Grade NATs
  - Cellular providers often use this technology
- Outage detection at the /24 granularity has limitations
- Systematic evaluation remains challenging due to incomplete ground truth

Contact Info

Alberto Dainotti

CAIDA, UC San Diego
alberto@caida.org
858-534-9249
@AlbertoDainotti