How to Find Correlated Internet Failures

Ramakrishna Padmanabhan[1,2(✉)], Aaron Schulman[3], Alberto Dainotti[2], Dave Levin[1], and Neil Spring[1]

[1] University of Maryland, College Park, USA
{dml,nspring}@cs.umd.edu
[2] CAID/UCSD, La Jolla, USA
{ramapad,alberto}@caida.org
[3] UCSD, San Diego, USA
schulman@cs.ucsd.edu

Abstract. Even as residential users increasingly rely upon the Internet, connectivity sometimes fails. Characterizing small-scale failures of last mile networks is essential to improving Internet reliability.

In this paper, we develop and evaluate an approach to detect Internet failure events that affect multiple users simultaneously using measurements from the Thunderping project. Thunderping probes addresses across the U.S. When the areas in which they are geo-located are affected by severe weather alerts. It detects a disruption event when an IP address ceases to respond to pings. In this paper, we focus on simultaneous disruptions of multiple addresses that are related to each other by geography and ISP, and thus are indicative of a shared cause. Using binomial testing, we detect groups of per-IP disruptions that are unlikely to have happened independently. We characterize these dependent disruption events and present results that challenge conventional wisdom on how such outages affect Internet address blocks.

1 Introduction

Even as residential users rely increasingly upon the Internet, last-mile infrastructure continues to be vulnerable to connectivity outages [1–3,5,18,20–24]. Measurement-driven approaches to study residential Internet failures will help improve reliability by identifying vulnerable networks and their challenges.

Techniques that detect outages at the Internet's edge often seek, using terminology from Richter et al. [19], *disruption events*: the abrupt loss of Internet connectivity of a substantial set of addresses. The set of addresses may comprise those belonging to the same /24 address block [18,19], BGP prefix [9], or country [4]. Techniques seek such disruption events because individually, each large disruption has impact and their size makes them easier to confirm, e.g., with operators. In contrast, disruptions affecting only a few users are harder to detect with confidence. For example, the lack of response from a single address might best be explained by a user switching off their home router—hardly an outage. However, residential Internet outages may be limited to a small neighborhood or apartment block; prior techniques are likely to miss such events.

D. Choffnes and M. Barcellos (Eds.): PAM 2019, LNCS 11419, pp. 210–227, 2019.
https://doi.org/10.1007/978-3-030-15986-3_14

In this work, we demonstrate a technique that detects disruption events with quantifiable confidence, by investigating the potential dependence between disruptions of multiple IP addresses in a principled way. We apply a simple statistical method to a large dataset of active probing measurements towards residential Internet users in the US. We find times when multiple addresses experience a disruption simultaneously such that they are unlikely to have occurred independently; we call the occurrence of such events *dependent disruptions*. Our preliminary results shed light on when, how large, and with which structure in the address space dependent disruptions happen. We show that even some large outages do not disrupt entire /24 address blocks.

Our contributions are:

- We demonstrate a technique to detect dependent disruption events using the binomial test.
- We show that dependent disruption events occur more frequently at night for some ISPs.
- The majority of dependent disruption events last less than an hour.
- We show that dependent disruption events do not always affect entire /24 address blocks and can therefore be missed by prior techniques that detect disruptions at this granularity [18,19].

2 Background and Related Work

In this section, we begin with a presentation of edge Internet disruption detection techniques. These techniques typically detect disruptions affecting a large group of addresses. Next, we provide a description of the Thunderping dataset [21] that yields per-IP address disruptions required for our detection technique.

2.1 Prior Work

Prior techniques that detect edge Internet disruptions typically detect disruptions that affect a group of addresses *collectively*. Like us, they also leverage the *dependence* among the per-IP address "disruptions" that these disruptions cause. However, they differ from our technique in that they look for dependence in large aggregates (that is, so many addresses are affected at the same time that there must be an evident anomaly) or limit their resolution to small address blocks, looking only for outages that cause dependent disruptions for most addresses in a monitored block.

Several systems investigate disruptions affecting a substantial set of addresses. The IODA system looks for the most impactful outages, those causing an extensive loss of connectivity for a geographical area or Autonomous System [4,7]. Hubble detects prefix-level unreachability problems [9] using a hybrid monitoring scheme that combines passive BGP monitoring and active probing.

Other systems detect disruptions affecting many addresses within /24 address blocks. For example, Trinocular uses historical data from the ISI census [6] to model the responsiveness of blocks and finds addresses within each block that are likely to respond to pings. The system pings a few of these addresses from each block at random in 11-minute rounds. It then employs Bayesian inference to reason about responses from blocks. When a block's responsiveness is lower than expected, Trinocular probes the block at a faster rate and eventually detects an outage when the follow-up probes also suggest the block's lack of Internet connectivity. Since Trinocular may not identify an outage even if a single address in a block responds to probing, it potentially neglects outages affecting /24 blocks only partially, including larger outages affecting multiple /24 blocks. Recently, Richter et al. used proprietary CDN logs to detect disruptions affecting multiple addresses within /24 address blocks [19]. They showed that many disruptions do not affect all addresses in a /24; we revisit this result in Sect. 4.4.

Disco [22] shares some features with our work: they also detect simultaneous disconnects of multiple RIPE Atlas probes within an ISP or geographic region to infer outages. However, there are two major differences between the Thunderping and RIPE Atlas datasets. At any given point in time, the Thunderping dataset typically consists of pings sent to thousands of addresses in relatively small geographical areas in the U.S. with active severe weather alerts. The Disco dataset consists of 10,000 RIPE Atlas probes distributed around the world; this sparse distribution may prevent the detection of smaller outages localized to one area (like a U.S. state). The second difference is that unlike Thunderping ping data whose timestamps are only accurate to minutes, the timestamps available in the RIPE Atlas datasets are accurate to seconds, permitting the use of Kleinberg's burst detection to detect bursts in probe disconnects.

2.2 The Thunderping Dataset Yields Per-Address Disruptions

The key insight behind our technique is that simultaneous disruptions of multiple individual IPv4 addresses could occur due to a common underlying cause. We therefore require per-IP address disruptions.

Such data is present in the Thunderping dataset [21]. Thunderping pings sampled IPv4 addresses from multiple ISPs in geographic areas in the United States. Originally designed to evaluate how weather affects Internet outages, the system uses Planetlab vantage points to ping 100 randomly sampled IPv4 addresses per ISP, from multiple ISPs, in each U.S. county with active weather alerts. Each address is pinged from multiple Planetlab vantage points (at least 3) every 11 min, and addresses in a county are pinged six hours before, during, and after a weather alert.

Here, we analyze a dataset of Thunderping's ping responses to detect disruptions for each probed address using Schulman and Spring's technique [21]. When an address that is responsive stops responding to pings from all vantage points that are currently probing it, we detect a disruption for that address. Since a disruption is detected only when all vantage points declare unreachability, the minimum duration of a disruption is 11 min (at the end of 11 min each vantage

point has pinged the address at least once).Thunderping continues to probe an address after it has become unresponsive, allowing us to estimate how long the unresponsive period lasted.

While per-IP address disruptions allow the detection of small disruption events, all per-address disruptions are not necessarily the result of Internet connectivity outages (e.g., a user might turn off their home router). This paper shows how to detect dependent disruption events using per-address disruptions.

3 Detecting Dependent Disruptions

In this section, we apply binomial testing to identify dependent disruptions in the outage dataset. First, we show how the binomial test works to rule out independent events and show how to apply the test to outages in reasonably sized aggregates of addresses. Second, we apply this method to the outage dataset, omitting addresses with excessive baseline loss rates and evaluating our chosen aggregation method. Finally we summarize the dependent disruptions we found in this dataset. This sets up analysis of these events (time of day, geography, and scope) which we defer to the following section.

3.1 Finding Dependent Events in an Address Aggregate

When many addresses experience a disruption simultaneously, there could be a common underlying cause. Such disruptions are statistically *dependent*. To identify these dependent events, our insight is to model address disruptions as *independent* events; when disruptions co-occur in greater numbers than the independent model can explain, the disruptions must be *dependent*. Binomial testing provides precisely this ability to find events that are highly unlikely to have occurred independently.

Given N addresses, the binomial distribution gives the probability that D of them were disrupted *independently* as:

$$\Pr[D \text{ independent failures}] \quad = \quad \binom{N}{D} \cdot P_d^D (1 - P_d)^{N-D} \tag{1}$$

where P_d represents the probability of disruption for the aggregate N. To apply this formula, we must first set a threshold probability below which we consider the simultaneous disruption to be too unlikely to be independent. We set this threshold to 0.01%. We then solve for D_{min}, the smallest (whole) number of simultaneous disruptions with a smaller than 0.01% chance of occurring independently. Table 1 in the appendix presents computed values of D_{min} for various values of N and P_d. This table shows that, even for large aggregates of IP addresses, often few simultaneous disruptions are necessary to be able to confidently conclude that a dependent disruption has occurred. As we will see, when applied to our dataset, D_{min} values are typically below 8.

There are two practical challenges in applying this test. First, we must choose aggregates of N IP addresses that define the scope of a dependent disruption:

too large an aggregate will have too large a chance of simultaneous independent failures and drive up D, while too small an aggregate may fail to include all the addresses in an event. Second, we must estimate P_d for each aggregate. We address each in turn.

3.1.1 Choosing Aggregate Sets of IP Addresses

Our technique assumes some *aggregate* set of IP addresses among which to detect a dependent disruption. We note that the *correctness* of our approach does not depend on how this set is chosen—the binomial test will apply so long as independent failures can be modeled by P_d. When applying our technique, IP addresses must be aggregated into sets that are large enough to span interesting disruption events, but not so large as to become insensitive to them.

In this paper, we aggregate IP addresses based on the U.S. state and the ASN they are in.*State-ASN* aggregates have the benefit of spanning multiple prefixes (so we can observe whether more than one /24 is affected by a given disruption event), but also being constrained to a common geographic region (so hosts in an aggregate are likely to share similar infrastructure). There are two limitations with this approach: states are not of uniform size, though the test elegantly handles varying N, and a few ISPs use multiple ASNs, which may hide some dependent failures. Alternate aggregations are possible (Appendix A.4).

3.1.2 Calculating the Probability of Disruption (P_d)

As a final consideration, we discuss how to estimate the probability of disruption, P_d, from an empirical dataset of disruptions. We assume that the dataset can be separated into a set of discrete "time bins"; this is common with ping-based outage detection, such as Thunderping and Trinocular, which both consider 11-minute bins of time. P_d can be estimated using the following equation:

$$P_d = \frac{\#\text{disruptions}}{\#\text{timebins}} \tag{2}$$

Here, #timebins represents the total number of observation intervals used: if a single host was measured across 10 time intervals and five other hosts were all measured across 3, then $\#\text{timebins} = 10 + 3 \cdot 5 = 25$.

We only consider state-ASN aggregates where we were able to obtain a statistically significant value for P_d. For statistical significance, we adhere to the following rule of thumb [25, Chap. 6]: we accept a state-ASN aggregate with t timebins and estimated probability of disruption P_d only if:

$$tP_d(1 - P_d) \geq 10 \tag{3}$$

3.2 Applying Our Method to the Thunderping Dataset

We investigate all ping responses in the Thunderping dataset from January 1, 2017 to December 31, 2017 and detect disruptions according to the methodology described above. During this time, Thunderping had sent at least 100 pings to 3,577,895 addresses and detected a total of 1,694,125 individual address disruptions affecting 1,193,812 unique addresses. The top ISPs whose addresses

Thunderping sampled most frequently include large cable providers (Comcast, Charter, Suddenlink), DSL providers (Windstream, Qwest, Centurytel), WISP providers (RISE Broadband), and satellite providers (Viasat). While most addresses have low loss rates, 2% of addresses had loss rates exceeding 10%; we remove these addresses to avoid biasing the analysis. We report additional details about these addresses in [15,17].

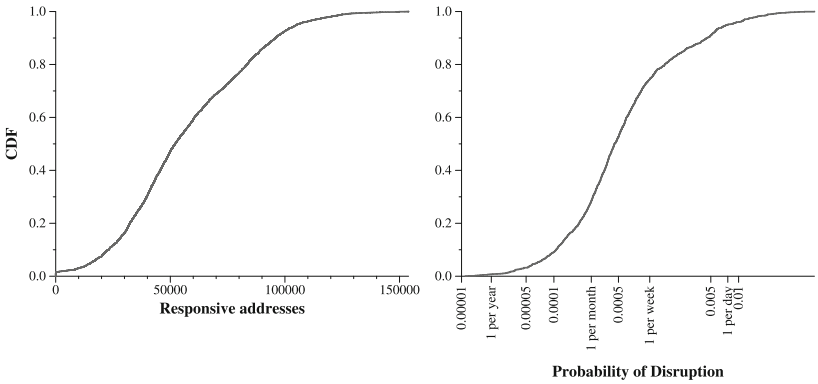


Fig. 1. Potential N and P_d values in the Thunderping dataset: on the left, we show the distribution of all addresses (across all state-ASN aggregates) pinged by Thunderping that can potentially fail in each 11 min time bin. On the right, we show the distribution of the probability of disruption (P_d) for the 1559 state-ASN address aggregates we studied.

Detecting Dependent Disruptions in the Thunderping Dataset

We use Fig. 1 to describe potential N and P_d values in the Thunderping dataset. On the left, we show the distribution of addresses pinged by Thunderping in each 11 min timebin in 2017. The median number is roughly 50,000 addresses across all U.S. states and ISPs. Since many weather alerts tend to be active at any given point of time, these addresses are likely to be distributed among tens of state-ASN aggregates. In 2017, the maximum addresses that could potentially fail in any state-ASN aggregate was 15,863. On the right, we show the distribution of P_d values for all state-ASN aggregates that we considered. There is extensive variation: addresses in some of these aggregates experience disruptions only once every year, whereas in other aggregates they experience disruptions more often than once per day.[1]

[1] Since disruptions are a superset of outages and dynamic reassignment [16], frequent disruptions are not necessarily indicative of poor Internet connectivity. Also, the existence of many aggregates with few disruptions indicates that Thunderping often pinged addresses during weather conditions that were not conducive to disruptions.

For each state-ASN aggregate, for each 11-min window during which Thunderping had pinged addresses, we identify the maximum number of addresses that can potentially fail, N, *i.e.,* all the addresses that are responsive to pings at the beginning of the window. Next, we apply the binomial test for each of these windows since we know N and P_d. When the number of disruptions in a window is at least D_{min}, we determine that a dependent disruption event occurred in that window with a probability greater than 0.9999.

In total, we detected 20,831 events with dependent disruptions in 2017. We analyzed our confidence in these dependent disruptions. The detailed results are included in the appendix (Fig. 8); in summary, the probabillity that detected events occurred independently is typically much smaller than our choice of 0.01%. We analyze the characteristics of these events next.

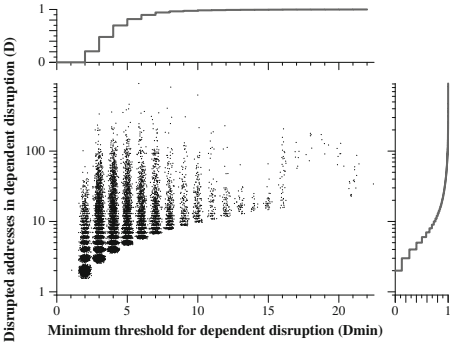


Fig. 2. For each detected dependent disruption event, Fig. 2 shows the D_{min} value on the x-axis and the corresponding number of observed disruptions on the y-axis. 62% of the 20,831 detected events had more than D_{min} observed disruptions. The scatterplot adds a random gaussian offset to both x and y with mean of 0.1, clamped at 0.45, to show density.

How Many Addresses Are Disrupted Dependently?

The binomial test does not say that *all* of the addresses that were observed to be disrupted during a dependent event were disrupted in a dependent manner. Consider if D_{min} is 4 and we detect an event where 7 addresses were disrupted. The binomial test shows us that the event took place with very low probability. However, that does not necessarily mean all 7 addresses were disrupted in a dependent manner; up to 3 of them ($D_{min} - 1$) could have been disrupted independently with up to 99.99% probability.

We call the set of addresses in a state-ASN aggregate that were disrupted in the time-bin of a dependent event the observed group of addresses that were disrupted, or the *observed disrupted group* for short. In the example above, the observed disrupted group contains 7 addresses. Of the observed disrupted group, our assumption is that some were disrupted together in a dependent manner:

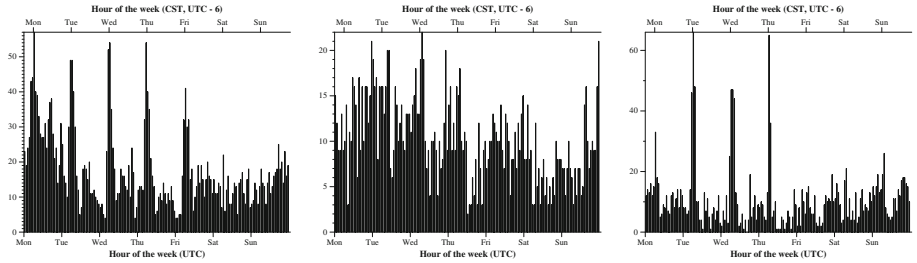
we call this subset the actual group of addresses that were disrupted, or *actual disrupted group*. We obtain a minimum bound on the actual disrupted group by subtracting $D_{min} - 1$ from the observed disrupted group; thus in the example above, the minimum number of addresses in the actual disrupted group is 4. For the 20,831 dependent disruption events, the total addresses in all the observed disrupted groups is 229,413 and the minimum total addresses in all the actual disrupted groups is 165,328.

We study the relationship between D_{min} for a state-ASN aggregate on the x-axis and the corresponding number of addresses in the observed group of disrupted addresses (on the y-axis) in Fig. 2. Each point corresponds to one of the 20,831 detected events. Sometimes, a state-ASN aggregate had such low P_d that even a single disruption in a 11-min bin occurred with less than 0.01% probably and therefore had a D_{min} value of 1. However, since we are looking for unlikely disruptions of multiple addresses, at least two addresses were disrupted in the same time-bin for all our detected events. For 12,911 (62%) detected events, *more* than D_{min} addresses experienced disruptions in the same time-bin, corroborating the result from Fig. 8 (in the appendix) that most detected events would have been detected even with a stricter threshold.

We detected dependent disruption events with various sizes as shown in Fig. 2. There are 693 (3%) events with more than 50 observed disrupted addresses. The largest detected event had 913 addresses experience disruptions in the same time-bin in AS33489 (Comcast) in Florida at 2017-09-13T20:33 UTC time. This detected event correlates to the minute with a known failure event for Comcast that was discussed in the Outages mailing list [14]. However, for most of the events, the size of the observed group of disrupted addresses is small: there were 2,593 (12%) with two, 2,969 (14%) with three, 2,776 (13%) with four, and 2,175 (10%) with five observed disrupted addresses. These results highlight the ability of our technique to detect even small sized disruptions with confidence.

4 Properties of Dependent Disruptions

In this section, we study various properties of dependent disruptions. For some properties, we conduct additional analyses on specific ISPs in the Thunderping dataset: Comcast (cable), Qwest (DSL) and Viasat (Satellite). These are three ISPs whose addresses are pinged frequently by Thunderping and where we were able to detect in excess of a thousand dependent disruption events (3109 events for Comcast, 1855 for Viasat, 1734 for Qwest). The appendix contains additional detail on per-ISP dependent disruption events.



(a) Comcast (b) Qwest (c) Viasat

Fig. 3. The y-axis shows dependent disruption events that began in each hour of the week. 'Mon' on the bottom x-axis refers to midnight on Monday in UTC time. On the top x-axis, 'Mon' refers to midnight at UTC-6 (CST).

4.1 Dependent Disruptions Are More Frequent at Night for Some ISPs

Richter et al. have recently shown that disruptions tend to happen more frequently during maintenance intervals close to midnight local time [19]. They did so using proprietary data from a CDN, collected at the granularity of every hour. Here, we investigate if our technique can identify similar patterns.

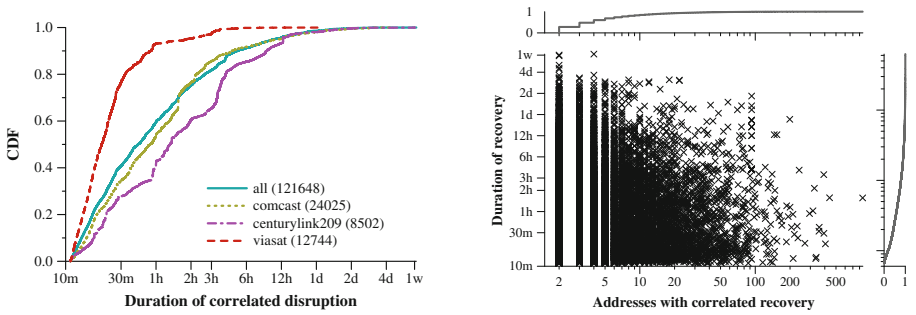
Figure 3 shows that individual ISPs can have different behavior. Comcast and Viasat have more dependent disruption events occurring close to midnight, CST, on weekday nights. Qwest, on the other hand, does not appear to have a clearly discernible pattern. Our results confirm those from prior work [19], lending credence to our technique.Moreover, we are able to do so using public (Thunderping) data and a granularity of every 11 min.

4.2 Dependent Disruptions Can Recover Together

Here, we investigate whether dependent disruption events are accompanied by *dependent recovery*. Since Thunderping continues to probe an IP address even after it becomes unresponsive (until six hours after the end of the weather alert [21]), it can observe when the address becomes responsive again. This responsiveness may signal that the disruption for the address has ended. Multiple addresses that are disrupted together and also recover together offer evidence that: (a) the event was indeed dependent and (b) the event has ended, allowing estimation of the disruption's duration.

Most dependent disruptions also have correlated recoveries. Of 20,831 dependent disruption events, 6,869 (33%) had *all* disrupted addresses recover during the same 11-min time-bin. Further, 14,789 (71%) disruption events had at least half of the disrupted addresses recover together. Across all of the 20,831 dependent disruption events, there were 229,413 observed disrupted addresses in total. Of these, 121,648 (53%) disrupted addresses—from 15,117 (73%) dis-

ruption events—exhibited a dependent recovery with other addresses from that same group. This indicates that dependent recovery is quite common.



(a) (b)

Fig. 4. (a) The distribution of durations of dependent disruptions for all addresses that recovered in a correlated manner. 60% of addresses recovered in less than an hour. (b) For dependent disruption events where at least two addresses recovered, this shows the number of addresses that recovered on the x-axis and the corresponding recovery duration for the event on the y-axis. Dependent disruption events vary in their duration irrespective of the number of affected addresses.

Recovery Times are Often Shorter than an Hour

Next, we turn our attention to the time it takes dependent disruptions to recover. Figure 4(a) shows that 60% of recovered addresses recovered in less than an hour. Our technique is able to identify this, because we operate at the precision of the 11-min time-bins from standard outage detection datasets. Conversely, recent work that finds disruptions spanning an entire calendar hour [19] would miss these disruptions.

Next, we examine whether short recovery durations can be attributable to small disruption events: that is, do the recoveries appear quick because only a couple hosts were disrupted? Figure 4(b) shows that the answer is no: Even dependent disruptions with hundreds of addresses that recovered together often last less than an hour.

4.3 Dependent Disruptions Can Be Multi-ISP

Dependent disruption events can also span multiple ISPs within a single state: these events indicate a fault of infrastructure shared by the ISP or their customers. Here, we broaden our analysis to examine whether dependent disruption events are correlated across multiple ISPs within the same state.

We observe 333 instances where multiple ISPs in the same state had simultaneous dependent disruption events, and we are able to confirm that many occurred on days when the media reported large power outages in those

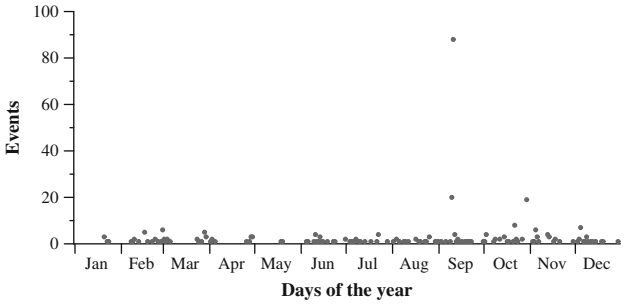


Fig. 5. Multi-ISP dependent disruption events over time: several ISPs in the same state have simultaneous disruption events on 333 occasions. Here, we show how many events occurred on each day of the year in 2017. Days with many multi-ISP events often correlate with days with large known power outages.

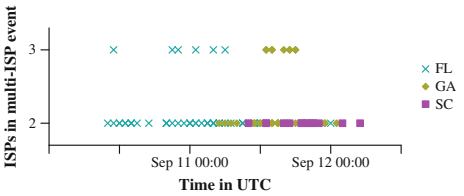


Fig. 6. Multi-ISP dependent disruption events during Hurricane Irma in Florida (FL), Georgia (GA), and South Carolina (SC). Of 111 events during this time, 15 affected 3 ISPs simultaneously and 96 affected 2.

areas. Figure 5 shows days in 2017 when multi-ISP dependent disruption events occurred. Of the 333 instances, 88 (26%) occurred on a single day during Hurricane Irma (Sep 11). Figure 6 shows multi-ISP events during Hurricane Irma by state and by the number of individual ISPs affected during each multi-ISP event. We observed 20 multi-ISP events in Florida on Sep 10, when Irma made landfall [8]. As Irma moved northwards, we saw multi-ISP events in Georgia and South Carolina as well. Other days with many such events include Oct 30 with 19 events across six states in the Northeastern U.S. (Maine, New Hampshire, Vermont, Connecticut, Massachusetts, Rhode Island); there were recorded power outages during this time as a result of a severe storm [11–13]. On Oct 22, there were 4 multi-ISP events in Oklahoma and 2 in Arkansas; there are corresponding reports of power outages during these times as well [10].

4.4 Dependent Disruptions May Not Disrupt Entire /24s

Here, we examine if dependent disruption events disrupt entire /24 address blocks. If so, they would likely be detected by prior work that looks for outages at these granularities [18,19]. If there continue to be responding addresses

within a /24 with a disrupted address, however, prior work may miss the disruption.

To analyze how dependent disruptions affect /24 address blocks, we find all addresses in the observed disrupted group for a dependent disruption event and group them by /24s. As a running example in this section, consider a dependent disruption event comprising 3 addresses in 1.2.3.0/24, 5 addresses in 2.3.4.0/24, and 2 addresses in 4.5.6.0/24. We call these the *observed disrupted* /24s for this event. For each of these /24s, we also find how many addresses were pinged by Thunderping that were responding to pings *before* the dependent disruption and that continued to respond for at least 30 min *after* the time-bin where the dependent disruption occurred. We term these addresses the responsive addresses in a /24 since these addresses were not affected by the disruption.

Our goal is to find how many /24 address blocks exist where at least one address within the /24 was an actual address in a dependent disruption but there were other addresses within the same /24 that continued to be responsive. Such /24s only experience a *partial* disruption (as defined in [19]). First, we checked how many of the 20,831 dependent disruption events had at least one responsive address in *all* of the observed disrupted /24s; there were 12,825 (61%) such events. For each of these events, even if some of the observed disrupted /24s for the event have addresses that failed independently, since all disrupted /24s continue to have at least one responsive address, all affected /24s only experienced partial disruptions (that could be missed by prior work).

Next, we investigate the subset of observed disrupted /24s where there were at least D_{min} failures within the /24 itself. Since the entire state-ASN aggregate only required D_{min} failures, when D_{min} or more addresses are disrupted within a single /24, the /24 has at least one actual disrupted address. We obtain the minimum bound on the number of actual disrupted addresses in a /24 by subtracting $D_{min} - 1$ from the observed disrupted addresses in that /24. Suppose the D_{min} for the example dependent disruption event above was 3. We would obtain a minimum bound of at least 1 actual disrupted address in 1.2.3.0/24. In 2.3.4.0/24, the minimum bound is 3. In 4.5.6.0/24, the minimum bound is 0 and we are unable to determine if the addresses in this /24 had a dependent disruption. Of 92,777 observed disrupted /24s (across all dependent disruption events), we find that 14,702 (16%) have at least D_{min} disrupted addresses. Each of these is a point in Fig. 7.

We find that many disrupted /24s with actual disrupted addresses have other addresses that continued to be responsive. 10,164 (69%) /24s had at least one responsive address, 9327 (63%) had at least two responsive addresses, and 6,096 (41%) had at least 10 responsive addresses. 1,691 /24s had at least 10 actual disrupted addresses; of those, 550 (33%) had at least 10 responsive addresses. In the appendix, we show that such behavior occurs across ISPs and we also discuss the implications of these results for prior work.

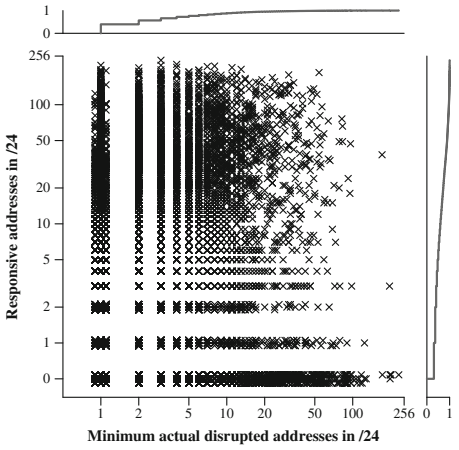


Fig. 7. Minimum actual disrupted addresses in a /24 vs. responsive addresses in a /24, for all /24s with at least D_{min} address that were disrupted during a detected dependent disruption event.

5 Discussion and Conclusion

We developed a technique to detect dependent disruption events with high confidence using the binomial test. The technique is general enough to apply to any dataset of disruptions; we applied it to the Thunderping dataset [21]. This dataset has the benefit of containing per-address disruption data from thousands of addresses in relatively small geographic regions (like a U.S. state) that may experience failures due to common underlying causes. We inherit the limitations of the Thunderping probing scheme—the system only measures residential addresses in one country (the U.S.), it probes mostly during times of predicted severe weather, and the minimum duration of disruptions is 11 min—so our conclusions may be limited in that they apply to this data.

Our application of the binomial test upon this dataset allowed us to show the feasibility of detecting large disruption events (such as power outages during times of severe thunderstorms) and also much smaller events. The majority of dependent disruptions last less than an hour although a small fraction continued for days. Corroborating prior work, we observe that disruption events occur more frequently at night for some ISPs. However, many disruptions do not affect entire /24 address blocks, suggesting that prior work may miss detecting them.

Simultaneous renumbering of entire prefixes by an ISP would manifest as a dependent disruption event. However, Richter et al. show that such events occur rarely in the U.S.; even elsewhere, they occur only in a few ASes [19]. Since Thunderping pings only U.S. addresses, the dependent disruption events we detected are unlikely to be caused by simultaneous renumbering. We believe that most of these events are caused by outages and are pursuing efforts to corroborate our inferences against ground truth.

Acknowledgments. We thank Arthur Berger, Philipp Richter, our shepherd Georgios Smaragdakis, and the anonymous reviewers for their thoughtful feedback. This research is supported by the U.S. Department of Homeland Security (DHS) Science and Technology Directorate, Cyber Security Division (DHS S&T/CSD) via contract number 70RSAT18CB0000015 and by NSF grants CNS-1619048 and CNS-1526635.

A Appendix

A.1 Determining D_{min}

Section 3.1 described our technique for detecting dependent disruptions through the calculation of D_{min}. Table 1 presents D_{min}, computed for various values of N and P_d. This table shows that, even for large aggregates of IP addresses, often few simultaneous disruptions are necessary to be able to confidently conclude that a dependent disruption has occurred.

Table 1. D_{min} values for varying values of N and P_d. There is less than 0.01% probability according to the binomial test that D_{min} or more addresses fail for each N and P_d.

N	D_{min}			
	P_d = 1/hour	1/day	1/week	1/month
10	8	3	2	2
50	21	5	3	2
100	35	7	4	3
500	126	14	6	4
1000	231	21	8	5
5000	1021	64	17	8
10000	1980	112	26	11
50000	9491	457	85	29

A.2 Analyzing the Confidence of Detected Disruption Events

Here, we examine our confidence in the 20,831 detected dependent disruption events from Sect. 3.2. The occurrence of D_{min} disruptions has less than 0.01% probability according to the binomial test. We test if most detected dependent disruption events have exactly 0.01% probability of occurring or if they are well clear of this threshold.

Figure 8 shows the distribution of the probability that we incorrectly classify an independent event as dependent. The probability of occurring independently is less than 0.005% for 90% of the events and less than 0.001% for 75%. Thus, the probabillity that detected events occurred independently is typically much smaller than our choice of 0.01%.

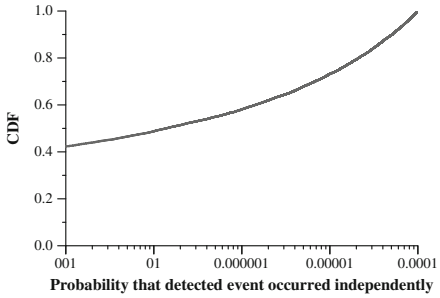


Fig. 8. Figure 8 shows the distribution of the probability that the 20,831 detected dependent disruption events could have occurred independently. For 90% of events, the probability of occurring independently is less than 0.00005.

A.3 Dependent Disruption Events Across ISPs

We grouped dependent disruption events by ISP to check if any ISPs contribute an unusual number of events. Figure 9 shows the top 15 ISPs with dependent disruption events. These top 15 ISPs together account for 13,643 (65%) of all detected events.

We emphasize that these results are not meant to reflect any underlying problems with these ISPs; Thunderping samples and pings large ISPs more frequently and consequently, finds more disrupted addresses in them. The purpose of this analysis is to ensure that no ISP contributes unduly many events.

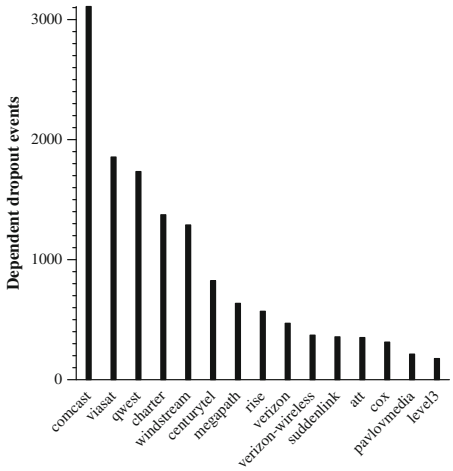
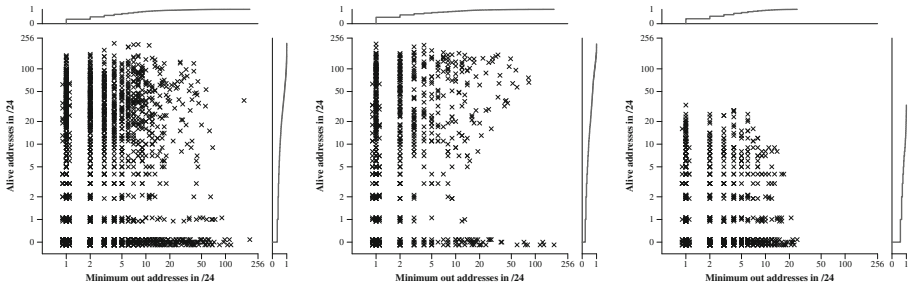


Fig. 9. Figure 9 shows the number of dependent disruption events detected per ISP. Note that these numbers are more a reflection of addresses sampled and pinged in the Thunderping dataset than any major underlying problem in their infrastructure. We leave per-ISP comparisons of dependent disruptions to future work.



(a) Comcast (b) Qwest (c) Viasat

Fig. 10. For Comcast, Qwest, and Viasat: Minimum actual disrupted addresses in a /24 vs. responsive addresses in a /24, for all /24s with at least D_{min} address that were disrupted during a detected dependent disruption event. All ISPs have /24s with actual disrupted addresses where there continued to be responsive addresses throughout the disruption.

A.4 Dependent Disruptions May Not Disrupt Entire /24s: Implications

Continuing our analysis from Sect. 4.4, we investigated if the responsiveness of other addresses in /24s with actual disrupted addresses would vary across ISPs. Figure 10 shows per-ISP behavior. We see that all these ISPs have /24s with actual disrupted addresses where there continued to be responsive addresses throughout the disruption.

Prior work detecting outages within /24 aggregates may miss these events. Since a single positive response from an address within a /24 could lead Trinocular to conclude that the block is alive [18], it can miss dependent disruption events affecting only a subset of addresses within a /24 address block. Richter et al.'s technique is capable of detecting partial /24 disruptions [19]; indeed, many of their disruptions did not affect all addresses in the /24. However, their choice of the *alpha* parameter in their technique ($alpha = 0.5$) meant that they would only detect disruptions where at least half of the active addresses were disrupted. In this paper, we showed that many /24s with actual disrupted addresses continued to have more than half of their (sampled) addresses responsive.

We believe that prior work may be able to detect these events by analyzing broader address aggregates (such as the state-ASN aggregates we use), in addition to /24 aggregates. In preliminary investigations, we found that many of our dependent disruption events consisted of multiple observed disrupted /24s that were each only partially disrupted; that is, a few addresses from many /24s were disrupted simultaneously but there continued to be other responsive addresses in these /24s. One of the largest events had 811 addresses from 42 /24 blocks in the observed disrupted group and 40 of these blocks had responsive addresses. We leave additional analyses for future work but we believe that we detected such events due to the broader aggregate of addresses we considered.

References

1. Argon, O., Bremler-Barr, A., Mokryn, O., Schirman, D., Shavitt, Y., Weinsberg, U.: On the dynamics of IP address allocation and availability of end-hosts. arXiv preprint arXiv:1011.2324 (2010)
2. Bischof, Z., Bustamante, F., Feamster, N.: The growing importance of being always on - a first look at the reliability of broadband internet access. In: Research Conference on Communications, Information and Internet Policy (TPRC), vol. 46 (2018)
3. Bischof, Z.S., Bustamante, F.E., Stanojevic, R.: Need, want. Broadband markets and the behavior of users. In: IMC, Can Afford (2014)
4. Dainotti, A., et al.: Analysis of country-wide Internet outages caused by censorship. In: IMC (2011)
5. Grover, S., et al.: Peeking behind the NAT: an empirical study of home networks. In: IMC (2013)
6. Heidemann, J., Pradkin, Y., Govindan, R., Papadopoulos, C., Bartlett, G., Bannister, J.: Census and survey of the visible Internet. In: IMC (2008)
7. Internet Outage Detection and Analysis (IODA). https://www.caida.org/projects/ioda/
8. National Hurricane Center Tropical Cyclone Report: Hurricane Irma. https://www.nhc.noaa.gov/data/tcr/AL112017_Irma.pdf
9. Katz-Basset, E., Madhyastha, H.V., John, J.P., Krishnamurthy, A., Wetherall, D., Anderson, T.: Studying black holes in the internet with Hubble. In: NSDI (2008)
10. Line Of Storms Moves Through Oklahoma. http://www.newson6.com/story/36651816/tornado-watch-in-effect-for-ne-oklahoma
11. Northeast Storm Undergoes Bombogenesis, Bringing 70 MPH Gusts, Almost 350 Reports of Wind Damage, Flooding—The Weather Channel. https://weather.com/forecast/regional/news/2017-10-30-northeast-storm-damaging-winds-flooding
12. 29–30 October 2017 damaging winds, heavy rainfall & flooding. https://www.weather.gov/aly/October29-302017
13. More than 1 million power outages in the Northeast after blockbuster fall storm - The Washington Post. https://www.washingtonpost.com/news/capital-weather-gang/wp/2017/10/30/over-one-million-power-outages-in-the-northeast-after-blockbuster-fall-storm/
14. Comcast outage on Sep 13 2017 in the Outages Mailing List. https://puck.nether.net/pipermail/outages/2017-September/010754.html
15. Padmanabhan, R.: Analyzing internet reliability remotely with probing-based techniques. Ph.D. thesis, University of Maryland (2018)
16. Padmanabhan, R., Dhamdhere, A., Aben, E., Claffy, K., Spring, N.: Reasons dynamic addresses change. In: IMC (2016)
17. Padmanabhan, R., Owen, P., Schulman, A., Spring, N.: Timeouts: beware surprisingly high delay. In: IMC (2015)
18. Quan, L., Heidemann, J., Pradkin, Y.: Trinocular: understanding internet reliability through adaptive probing. In: SIGCOMM (2013)
19. Richter, P., Padmanabhan, R., Plonka, D., Berger, A., Clark, D.: Advancing the art of internet edge outage detection. In: IMC (2018)
20. Sánchez, M.A., et al.: Dasu: pushing experiments to the internet's edge. In: NSDI (2013)
21. Schulman, A., Spring, N.: Pingin' in the rain. In: IMC (2011)
22. Shah, A., Fontugne, R., Aben, E., Pelsser, C., Bush, R.: Disco: fast, good, and cheap outage detection. In: TMA (2017)

23. Shavitt, Y., Shir, E.: DIMES: let the internet measure itself. SIGCOMM Comput. Commun. Rev. **35**, 71–74 (2005)
24. Sundaresan, S., Burnett, S., Feamster, N., de Donato, W.: BISmark: a testbed for deploying measurements and applications in broadband access networks. In: USENIX ATC, June 2014
25. van Belle, G., Heagerty, P.J., Fischer, L.D., Lumley, T.S.: Biostatistics: A Methodology for the Health Sciences, 2nd edn. Wiley, Hoboken (2004)