Improving the Optics of Active Outage Detection (extended)

USC/ISI Technical Report ISI-TR-733

May 2019

Guillermo Baltra
ISI/USC
baltra@isi.edu

John Heidemann
ISI/USC
johnh@isi.edu

ABSTRACT

There is a growing interest in carefully observing the reliability of the Internet's edge. Outage information can inform our understanding of Internet reliability and planning, and it can help guide operations. Outage detection algorithms using active probing from third parties have been shown to be accurate for most of the Internet, but inaccurate for blocks that are sparsely occupied. Our contributions include a definition of outages, which we use to determine how many independent observers are required to determine global outages. We propose a new *Full Block Scanning* (FBS) algorithm that gathers more information for sparse blocks to reduce false outage reports. We also propose *ISP Availability Sensing* (IAS) to detect maintenance activity using only external information. We study a year of outage data and show that FBS has a True Positive Rate of 86%, and show that IAS detects maintenance events in a large U.S. ISP.

1 INTRODUCTION

Internet reliability is of concern to all Internet users, and improving reliability is the goal of industry and governments. Yet government intervention, operational misconfiguration, natural disasters, and even regular weather all cause network outages that affect many. The challenge of measuring outages has prompted a number of approaches, including active measurements of weather-related behavior [12], passive observation of government interference [3], active measurement of most of the IPv4 Internet [9], passive observation from distributed probes [13], and analysis of CDN traffic [11].

Each outage measurement system has strengths and weaknesses. Passive approaches and CDN analysis provide unique insight into firewalled areas that will not respond to active probing, but require longer observation and so provide somewhat less temporal or spatial precision. Active systems cannot see behind firewalls or NAT, but allow control of the precision and coverage and can provide very broad coverage. This control allows active systems to provide relatively precise timing of outage onset and duration (typically 11 minutes, compared to 60 minutes or more for passive), and scope (individual IP addresses or /24 blocks of addresses), but the cost is measurement traffic that is sometimes misinterpreted as malicious, drawing complaints or firewalling. Thus a fundamental trade-off in active systems is balancing correctness and precision with minimizing measurement traffic.

A huge advantage of multiple outage measurement systems is their results can be compared. Recent work in CDN-based analysis has provided confidence in the field, showing most blocks show similar responses with both active measurement and CDN analysis [11]. However, it also showed that a few percent of blocks are challenging to measure, generating frequent outage reports that are false positives—not confirmed in passive data. These relatively few blocks can generate *many* incorrect outage reports, so while active finds 94% of passive outage events, passive confirms only 74% of active events. This work highlights two sources of these differences: a few blocks where active outage detection makes premature decisions, and ISP maintenance activity.

As the field matures, a second problem is that we need a definition of network outages that is independent of any measurement system. Such a first-principles definition will allow us to work towards evaluating how closely different outage detection systems compare to a "platonic ideal". It also allows us to answer pressing operational questions, such as how many observers are required to avoid bias.

This paper makes three contributions. First, we propose a theoretical definition of an Internet outage (§3.1). We then use this definition and data to show, empirically, that *three independent observers* are sufficient to identify global outages with active detection (§4.1).

Second, we develop *Full Block Scanning* (FBS), an algorithm to improve outage detection in blocks where use is sparse or transient (§3.2). Examination of random down events show FBS has a True Positive Rate of 86% (§4.2).

Our final contribution is the *ISP Availability Sensing* (IAS) algorithm. IAS detects ISP maintenance with only external information (§3.3). We evaluate it with a natural experiment at CenturyLink, showing that IAS corrects 23 events, 13 confirmed in their maintenance window (§4.3).

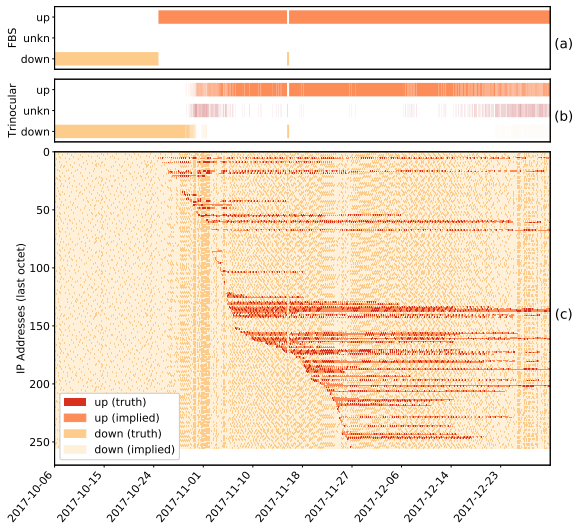

Figure 1: A sample block over time (columns). The bottom (c) shows individual address as rows, with colored dots when the address responds. The middle bar (b) shows Trinocular status (up, unknown, and down), and bar (a) is Trinocular with Full Block Scanning.

All of the data used in and created this paper will be available at no cost. Our work was IRB reviewed and identified as non-human subjects research (USC IRB IIR00001648). .

2 CHALLENGES TO ACTIVE DETECTION

Our goal is to detect outages with third-party, active measurements. Such approaches provide large coverage (more than 3M blocks [9]), but we must address several causes of false outages: partial outages, sparse blocks, and usage changes.

2.1 Problem: Partial Outages

A partial outage is a destination that is not reachable from one location but is reachable from elsewhere on the Internet. They may occur due to link or route failures in the middle of the path. Hubble states that two-thirds of reachability problems are what are partial outages [5].

Figure 1 shows an 8.5 hour-long outage on 2017-11-16. This event is an outage because we see the same addresses (rows) active (dark dots) before and after the event, but no addresses are active during it. We know this outage is *local* to this Vantage Point (VP) because other VPs (not shown) can reach these addresses the whole time.

Trinocular handles partial outage by reporting the target block up if any VP can reach it [9]. In §4.1 we explore alternative voting strategies and how many VPs are needed.

2.2 Problem: Sparse Blocks

Sparse blocks pose a problem in that active scanning will find many empty addresses. When probing is rate-limited, these non-responses may result an a hasty decision that marks the block down, a *false outage*.

To constrain traffic to each block, and to track millions of blocks, Trinocular limits itself to 15 probes per round. This limit can cause incorrect decisions two ways. First,it may fail to reach a definitive belief and mark the block *unknown*. Alternatively, if the block is usually responsive, several non-responses may produce a down belief.

In Figure 1(b), the middle bar shows that Trinocular often marks the block unknown for the week starting 2017-10-30, and again for weeks after 2017-12-12. Trinocular knows all addresses in this block *may* be used (all addresses are in $E(b)$, defined as the set of ever-active addresses), but in both of these periods, only a few *are* used—the block is sparse ($A(E(b))$, defined as the probability $E(b)$ responds, is small).

Prior work has filtered sparse blocks with several ad hoc ways: Trinocular marks very sparse blocks as *unmeasurable* (when $A(E(b)) < 0.10$ or $|E(b)| < 15$), and dropped blocks when observed A doesn't match predicted A [9]. Richter's use of Trinocular data dropped all blocks with more than 5 outages in 3 months [11], based on our recommendation. Trinocular notes that its unmeasurability test is not strict enough: indeterminate belief can occur when the $A(E(b)) < 0.3$ and $|E(b)| \geq 15$. Moreover, later work [10] dynamically tracked A, but Richter et al. showed that block usage changes dramatically, so blocks can *become* overly sparse.

We can define the sparseness of the block by its current availability ($\hat{A}_s(E(b))$. We consider blocks sparse when it is less than a threshold ($\hat{A}_s(E(b)) < T_{sparse}$), currently with T_{sparse} of 0.2. We observe that blocks show frequent outages (like Figure 1) when they are sparse. We find that 85% of blocks with 10 or more outages meet the sparse threshold, and yet sparse blocks represent only 22% of all blocks. (Graphs showing these results are omitted due to space.)

2.3 Problem: Changes in Usage

Outage detection systems determine a block is reachable because it sends traffic, responding to probes for active detection, or for other reasons with passive detection. However, traffic may be absent for reasons other than an outage: a network operator may reassign dynamic addresses, shifting users to other blocks, or users may deploy firewalls. Changes in block usage may cause false outages.

We see an example of usage change in Figure 1. The block has no activity for three weeks, then sparse use for a week, then moderate use, and back to sparse use for the last two weeks. Reverse DNS suggests this block uses DHCP, and gradual changes in use suggest the ISP is migrating users. The block was provably reachable after the first three weeks.

Before then it may have been reachable but unused, a false outage because the block is inactive.

3 DESIGN

We next describe our definition of outages and algorithms to correct two sources of false outages.

3.1 Outages: Global and Partial

We first *define* global outages so we can consider how to detect partial outages, proposing the following theoretical definition: **A *network outage* is the *unavailability* of an *active* network from *more than half* of the *Internet*.** By unavailable we mean lack of a logical or physical path to or from a network, or lack of response from a host. By active network we mean some prefix of addresses that is likely to fail together, and that is intended to be on the public Internet. By Internet we mean as observed from all public IP addresses. This definition implies that unavailability from less than 50% of the Internet is a *partial* outage.

Translating this theory to practice requires care, since measurement systems observe from a few (or at most, tens) of VPs. (But three independent sites are enough, see §4.1.)

We also must recognize that down events that *appears* to be an outage may be false, because only the network operator knows what networks are *active*. Richter et al. used internal information from clients to demonstrate that address reassignment cause false outages, defining *disruptions* to include both true and false outages [11]. We show a new algorithms to identify disruptions with only external information in §3.3, and define *up* and *down* events as the raw observations.

3.2 Full Block Scanning for Sparse Blocks

The challenge of evaluating sparse blocks is that Trinocular makes decisions on too little information, forcing a decision after 15 probes (a *Trinocular Round* or *TR*, every 11 minutes), even without reaching a definitive belief. We address this problem with *more* information: we *Scan the Full Block* (in a *Full Round* or *FR*) to consider all addresses that have ever been active (all of $E(b)$). By requiring decisions to consider all addresses we overcome the 15-probe limit. For Trinocular, we maintain this limit per Trinocular Round, but we assemble Full Rounds from "enough" adjacent Trinocular Rounds.

Formally, we define a Full Round ending at time t as the minimum N TRs before t that cover all $E(b)$ ever-active addresses of the block: $\sum_{i=t-N}^{t}(|TR_i|) \geq |E(b)|$.

Full Block Scanning (FBS) tests outages from Trinocular. If the block is currently sparse ($\hat{A}_s < T_{sparse}$) and the most recent Full Round included a positive response, then we override the outage. That is, if there are any positive responses in the last Full Round FR_t, we convert any outages to up if $\forall TR_i$ where $i \in [t-N, t]$.

We only use FBS for blocks where it is required because they are recently sparse. A block is recently sparse if the short-term running average of the response rate for the block $\hat{A}_s^{3FR}$, computed over the last three *FR*s, is below the sparse threshold ($\hat{A}_s^{3FR} < T_{sparse}$).

Finally, since FBS uses information from several TRs, it decreases the temporal precision of outage reporting. We know the block was up at the last positive response, and we know it is down after the full round of non-responses, so an outage could have begun any time in between. We therefore select a start time as the time of the last confirmed down event (the first known lit address, now down). That time has uncertainty of the difference between the earliest possible start time and the confirmed start time.

3.3 ISP Availability Sensing for Maintenance

We next identify disruptions due to ISP maintenance events that cause false outages due to inactive blocks. ISP maintenance actions such as renumbering users to different IP addresses appear as outages to an external observer when active blocks become inactive. However, they are disruptions and not user-affecting outages, since the users continue to get service at new IP addresses. Richter et al. identified this problem with data from sensors *inside* the ISP; we next identify such disruptions using *only external* information.

Our insight is that in user-affecting outages, ISPs *lose* active addresses, but maintenance with renumbering merely reassigns users to different addresses. We therefore detect non-outage disruptions with *ISP Availability Sensing* based on stability of the number of active addresses across the ISP.

Our approach starts by tracking the availability (the total number of active addresses) in the ISP. The availability of AS a at time t is $\mathcal{A}_t(a) = \sum_{b \in B_a} \hat{A}_s^{1FR}(b)$, where $\hat{A}_s^{1FR}(b)$ is the availability of block b over the last Full Round before time t, and B_a are all the blocks in AS a.

When the number of active addresses is roughly unchanged we infer that block-level outages are renumbering events (since each address departure is matched by another address appearance). We look for change Δ_t, defined as the difference in availability estimates between the last two adjacent Full Rounds (t and t'): $\Delta_t = (\mathcal{A}_t - \mathcal{A}_{t'})/(\max(\mathcal{A}_t, \mathcal{A}_{t'}))$. We compare to a threshold, so an event is maintenance if $\Delta_t \leq T_u$. We currently set $T_u = 0.05$, meaning real outages require a 5% drop in addresses.

4 EVALUATION

We next evaluate correctness of our three algorithms: partial, outages, sparse blocks, and block usage change.

4.1 How Many Vantage Points Is Enough?

We first consider outage definition (§3.1 to explore how many vantage points we need to rule out partial outages. Recall that a partial outage (an outage near a vantage point) makes it seem to that VP that most of the Internet is down. Partial outages can be removed by voting from independent

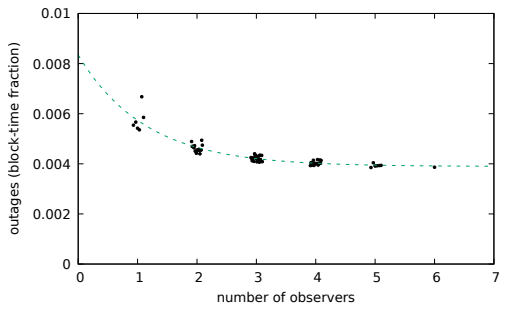

Figure 2: Down time fraction for all 63 possible combinations out of six sites with jittered individual readings. Dataset: A28.

observers. We next evaluate how many VPs are needed, how to confirm independence, and choices for how to vote.

4.1.1 Converging on Global Outages. When sampling from a distribution, additional samples converge on the true value as they reduce the margin of error. (For a normal distribution, confidence intervals on the mean fall as the square-root of the number of samples.) Although we do not posit a distribution for Internet outages, we expect to see diminishing changes as new vantage points to eliminate partial outages, and we converge on global outages.

To evaluate this question we take observations from a large number of vantage points and then look at subsets to see if they converge. Assuming VPs are basically independent (evaluated in §4.1.2), we can explore 63 possible combinations and test empirical convergence. Here we merge with any-up voting; we revisit voting in §4.1.3.

Figure 2 shows the fraction of time blocks are out, observed over 3 months of data for dataset A28 in 2017q2. Each point is one of the 63 possible combinations of 6 VPs (6 with 1 VP, 15 combinations of 2 VPs, etc., to 1 with all 6). For each cluster, we jitter points in the x axis to avoid plotting combinations on top of each other. (For this analysis we do not use the FBS and IAS algorithms to focus on basic detection. Use of those algorithms would flatten curve slightly.)

This data confirms our hypothesis of diminishing returns. Visually, the observations converge on 0.0039. Quantitatively, an exponential of $(0.0043x^{-0.875} + 0.0039)$ fits the data well, with a correlation coefficient of 0.9371.

This data suggests that 4 sites are definitely enough to filter partial outages (the worst combination of 4 sites is 0.0042, within 7%). In fact, we claim that **three independent sites are a reasonable good match to global outages**, with 12% (0.0044) of the convergence. While we believe three sites are sufficient and four sites are safe, of course more sites do not hurt convergence, so we recommend using more if available.

Table 1: *(a)* Similarities between sites relative to all six. Dataset: A33, 2018q3. *(b)* Comparison of different voting policies with 6 VPs. Dataset: A28, 2017q2.

	j	e	c	g	n
w	0.079	0.064	0.084	0.061	0.093
j		0.061	0.168	0.154	0.118
e			0.069	0.068	0.073
c				0.128	0.075
g					0.130

(a)

policy	up	down
any up	0.9961	0.0039
two up	0.9960	0.0040
1/2 up	0.9956	0.0044
2/3 up	0.9951	0.0049
all−{1}	0.9917	0.0083
all up	0.9174	0.0826

(b)

4.1.2 Are the sites correlated? Our evaluation of convergence assumes VPs do not share common network paths. VPs in different physical locations must have dissimilar access links, and with the trend towards a "flatter" Internet graph [6], they should have diverse WAN paths. We next quantify this similarity to validate our assumption.

We next measure the similarity of observations between each pair of VPs. We examine only cases where one of the pair disagrees with some other VP, since when all agree, we have no new information. If the pair agree with each other, but not some other VP, the pair show similarity. If they disagree with each other, they are dissimilar. We quantify similarity S_P for a pair of sites P as $S_P = (P_1 + P_0)/(P_1 + P_0 + D_*)$, where P_s indicates the pair of sites agree on the network having state s of up (1) or down (0) and disagree with the others, and for D_*, the pair disagree with each other. S_P ranges from 1, where the pair always agree, to 0, where they always disagree.

Table 1(a) shows similarity values for each pair of the 6 Trinocular VPs (locations: w: Los Angeles, j: Tokyo, e: Washington, DC, c: Colorado, g: Athens, n: Amsterdam). (We show only half of the symmetric matrix.) No two sites have a similarity more than than 0.17, and many pairs are around 0.08. This result shows that no two sites are particularly corrected, although j, g, and n (the three non-U.S. sites) seem more correlated than the others.

4.1.3 Voting Options To Resolve VP Disagreement. Partial outages introduce conflicting observations between VPs. Our definition of outages in §3.1 suggests majority voting defines truth, but can we reduce the billions of theoretical observers to just a few VPs?

We evaluate this question by comparing six voting strategies: any-up, 1/3-up, 1/2-up, 2/3-up, all-but-one up, and all-up. *Any-up* uses the intuition that if any VP can reach a target, it is up. Any-up produces false positives when the target block and the single vantage point observing it as active are both behind the failed router. For example, a VP inside a university in Athens would force that university's blocks up, even if their connection to the Internet was down. Other

strategies are less strict. Since networks are usually up, we break ties in favor of "up-ness".

Table 1(b) shows up and down fractions with 6 VPs for each voting strategy, for 3 months (A28 in 2017q2). (Because sites sometimes fail over the quarter, voting occasionally uses fewer than six sites.) The any-up policy has the lowest fraction of down time, and less strict policies result in lower rates of outages. The all-up policy is overly strict, since an outage local to *any* VP will break the vote and suggest that the target block is down. The other policies are all similar, although all-{1} shows some sensitivity failures at two VPs.

We conclude that *majority voting (1/2-up) is a good choice*, even with only 6 VPs.

4.2 Full Block Scanning Reduces Noise?

We next examine Full Block Scanning accuracy (§3.2) through studies of one block and then a country.

4.2.1 Case Study of One Block. Figure 1 shows one block with outage analysis as a case study. This block is in CenturyLink (AS209, a U.S. ISP).

This block is initially inactive (before 2017-10-24; the left 20% of the plot), then becomes sparsely used and relatively densely used (around 2017-11-01), but back to sparsely used (at 2017-12-23). It shows a partial outage on 2017-11-16.

Trinocular results ((b), the middle graph) show frequent unknown states that result in false outages, particularly when block usage is sparse in October and late December.

By contrast, Full Block Scanning ((a), the top graph), resolves this uncertainty. FBS' more information confirms the block is usually up, while recognizing the partial outage.

4.2.2 True Positives: Is Full Block Scanning Successful at Removing Noise? From this single block example, we next consider a country's Internet. Our goal is to see if FBS reduces noise by examining "true positives" (blocks correctly recovered by FBS because they were observation noise).

We study series of known outages that affected Iraq in February 2017. That country had 7 government-mandated Internet outages (the local mornings on February 2, and also the 4th through 9th) with the goal of preventing cheating during academic placement exams [4]. We identified 1176 Iraqi blocks using Maxmind's city-level database [7], and examined outages from 2017q1 Trinocular (dataset A27 [14]).

Figure 3 shows the number of Iraqi blocks that are detected as out during 2017q1, grouped into in 4096s timebins. We show outages without Full Block Scanning (the darker blue, top line) and with it (the lighter green line). The Iraqi exam week is highlighted in gray on the left, and we plot that week with a larger scale on the right.

In each of the 7 large peaks during exam week, most Iraqi blocks (nearly 1000) are out—our true outages. We also see that few blocks (5 to 15) are often down, likely false outages.

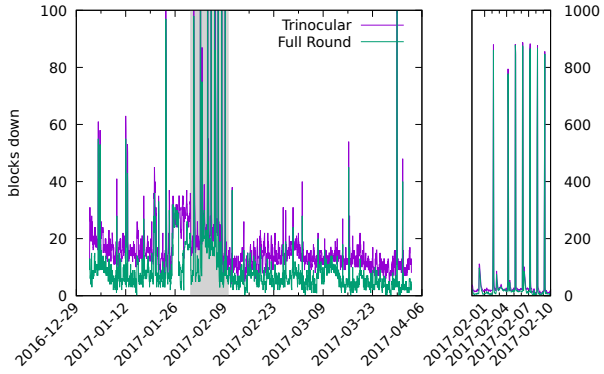

Figure 3: Down events in Iraqi blocks during 2017q1. Dataset A27.

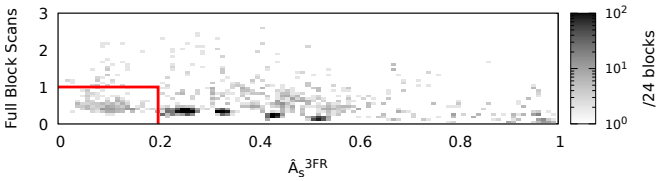

Figure 4: Outage events during the 7 Iraqi outages, measured of their $\hat{A}_s^{3FR}$ and full round values. Dataset: A27 (2017q1) subsetted to the 7 outage periods.

We see that FBS reduces the number of down blocks by 5–10 per observation interval, cutting the background noise of outages by about 36%. We confirm this reduction was due to noise by examining blocks marked down by Trinocular and up with FBS in 10 randomly-selected time periods. Of the 34 total down events, 33 (97%) were in sparse blocks that resemble Figure 1; the other block was diurnal. This study confirms that FBS recovers false outages due to sparseness.

4.2.3 False Positives: Does FBS Remove Legitimate Outages? We next look at how Full Block Scanning interacts with known outage events. Its goal is to remove noise and false outages, but if it is too aggressive it may accidentally remove legitimate outages (a "false positive" detection).

We look for false positives in the 2017q1 Iraqi exam dataset (Figure 4 in §4.2.2), using the 7 nationwide outages as ground truth and comparing Trinocular with and without FBS.

The 7 peaks in Figure 3 show known Iraqi outages, and we see little visual difference. When we examine the peaks of outages, FBS removes about 12 of the 880 blocks (about a 1.5% drop in outages). We believe these results represent an acceptable tradeoff in light of consistent reduction of noise during regular operation (Figure 4).

To confirm which outage events are filtered by FBS during the Iraqi outages, we identified all outage events during the

7 periods of Iraqi outages (from times 8am to 10am (local Iraqi time) on days 2017-02-02 and -10. Each outage event represents one block that was down for about 2 hours. We evaluate the $\hat{A}_s^{3FR}$ value and the duration of the outage event in Full Rounds, and show that scatter plot in Figure 4.

FBS filters outage events of sparse blocks ($\hat{A}^{3FR} < 0.2$) that last one Full Round (highlighted by the red box in the lower, left corner). We see that most outage events (93%) fall outside this region, either because they last longer or are in blocks that are more full. Outage events within the region represent a 7% of the total events in the figure. This analysis confirms why FBS successfully passes through the Iraqi outages with little change, but still is able to filter noise.

4.2.4 Random Sampling of Outage Events. Finally, we confirm our results with a random sample of events.

We draw 50 random blocks that show outages in Trinocular without FBS. We examine each block and their 5200 down events to determine best-effort ground truth. Of the 5200 down events, we see 4133 True Positives (79% of outages in sparse blocks are fixed by FBS), 621 false negatives (12% are not fixed), and 446 true negatives (9% true outages are not changed). The result is a True Positive Rate of 0.86 (621 false positives of 4752 true cases), so it is reasonably successful at removing noise. Many of the false negatives are due to moderate sparse blocks ($0.2 < A(E(b)) < 0.4$).

4.3 Does ISP Availability Sensing Handle Address Reallocation?

We next examine how well ISP Availability Sensing (IAS, §3.3) handles maintenance activity by network providers.

4.3.1 True Positives: Does IAS Detect Maintenance? We evaluate if IAS detects maintenance activity in 2017q4 with CenturyLink (AS209), a large U.S. ISP and transit provider. They have public maintenance windows (early morning Sunday, Tuesday and Thursday) [1], and report specific events [2]. Since we know of no public, inside-ISP data (like [11]), we use this natural experiment to test IAS. We identify CenturyLink blocks from 18 peers in Routeviews [8].

Figure 5 shows one quarter of data (2017q4). The bottom shows Δ_t, changes in active addresses, from the L.A. Trinocular instance. The top shows down events from 6-site merged results after FBS, but without IAS. We find 23 events in this merged data, each involving 35 or more blocks. Of the 23 events, more than half (13, indicated with capital letters) are in CenturyLink's published maintenance window. Five of these (indicated with *) correspond to events in the service log on their website. IAS identifies *all* these events as maintenance, except for event (o). These events are true positives.

Event (o) on 2017-11-16 (in red) is very large, and IAS passes it through as an actual outage when observed from

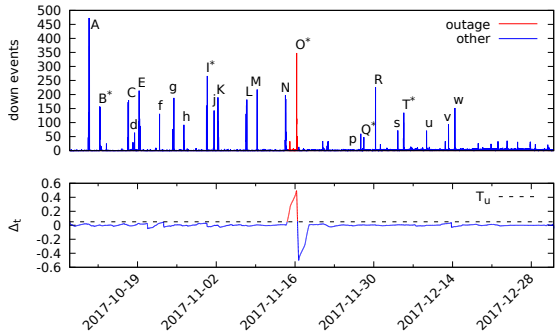

Figure 5: Down events (top) from six observers (top) and Δ_t (bottom) from Los Angeles. CenturyLink AS209. Dataset: A30, 2017q4.

this VP. This event is unusual, in that it was much larger (affecting 20,211 blocks, more than half) and longer (8.5 hours) at this VP in Los Angeles than from other sites (where it was 348 blocks and 2 hours). We believe this event was a local problem affecting Los Angeles. It shows that the IAS will correctly pass through large outages (a true negative).

4.3.2 False Positives: Does IAS Remove Legitimate Outages? Beyond the partial outage just described (a true negative), we next look for false positives in IAS.

We again use the Iraqi exam outages (§4.2.2). We apply IAS to the largest Iraqi ISP, AS50710, EarthLink Ltd. Communications & Internet Service. (This ISP is different from AS3703, earthlink.net, the large U.S. ISP purchased by Windstream Communications in March 2017.)

Of the seven exam outages, IAS passes the first four unchanged (a true negative). However, it removes some blocks from the three events, losing 19% of down events overall. We believe these false positives are because closely spaced true outages distort Δ_t; ongoing work considers adaptive thresholding that may lose fewer events.

5 RELATED WORK

Several groups have different methods to detect outages at the Internet's edge: ThunderPing first used active measurements to track weather-related outages on the Internet [12]. Dainotti et al. use passive observations at network telescope to detect disasters and government censorship [3], providing the first view into firewalled networks. Trinocular uses active probes to study about 4M, /24-block level outages [9], the largest active coverage. Disco observes connectivity from devices at home [13], providing strong ground truth, but limited coverage. Finally, Richter et al. detect outages with CDN-traffic, confirming with software at the edge [11]. They define disruptions, showing renumbering and frequent disagreements in a few blocks are false positives in prior work. Our work builds on prior active probing systems and the

Trinocular data and algorithms, and addresses problems identified by Richter while using only external information.

In addition to Richter et al.'s definition of disruptions and non-outage events [11], the early Hubble work defined partial outages and estimated about two-thirds of reachability problems are partial [5]. While prior work has had operational definitions of outages and disruption, we provide a theoretical outage definition that we believe applies generally.

6 CONCLUSIONS

This paper makes three contributions: an outage definition that helps identify the required number of observers, the FBS algorithm to reduce noise from active measurement of sparse blocks, and the IAS algorithm to detect ISP maintenance with only external observations. We showed these algorithms work well using multiple datasets and natural experiments; they can improve existing and future outage datasets.

ACKNOWLEDGMENTS

We thank Yuri Pradkin for his input on the algorithms and paper.

The work is supported in part by the National Science Foundation, CISE Directorate, award CNS-1806785; by the Department of Homeland Security (DHS) Science and Technology Directorate, Cyber Security Division (DHS S&T/CSD) via contract number 70RSAT18CB0000014; and by by Air Force Research Laboratory under agreement number FA8750-18-2-0280. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon.

REFERENCES

[1] CenturyLink. 2016. CenturyLink IQ Retail Service Level Agreement. (Jan 2016). http://www.centurylink.com/legal/docs/CenturyLink-IQ-SLA.pdf

[2] CenturyLink. 2019. Event History. https://status.ctl.io/history. (2019).

[3] Alberto Dainotti, Claudio Squarcella, Emile Aben, Marco Chiesa, Kimberly C. Claffy, Michele Russo, and Antonio Pescapé. 2011. Analysis of Country-wide Internet Outages Caused by Censorship. In *Proceedings of the ACM Internet Measurement Conference*. ACM, Berlin, Germany, 1–18. https://doi.org/10.1145/2068816.2068818

[4] Doug Madory. 2017. Iraq Downs Internet To Combat Cheating...Again! https://dyn.com/blog/iraq-downs-internet-to-combat-cheating-again/. (2017). Accessed: 2019-01-08.

[5] Ethan Katz-Bassett, Harsha V Madhyastha, John P John, Arvind Krishnamurthy, David Wetherall, and Thomas E Anderson. 2008. Studying Black Holes in the Internet with Hubble. In *Proceedings of the USENIX Conference on Networked Systems Design and Implementation*. ACM, San Francisco, CA, 247–262.

[6] Craig Labovitz, Scott Iekel-Johnson, Danny McPherson, Jon Oberheide, and Farnam Jahanian. 2010. Internet Inter-Domain Traffic. In *Proceedings of the ACM SIGCOMM Conference*. ACM, New Delhi, India, 75–86. https://doi.org/10.1145/1851182.1851194

[7] MaxMind. 2017. GeoIP Geolocation Products. http://www.maxmind.com/en/city. (2017).

[8] D. Meyer. 2018. University of Oregon Routeviews. http://www.routeviews.org. (2018).

[9] Lin Quan, John Heidemann, and Yuri Pradkin. 2013. Trinocular: Understanding Internet Reliability Through Adaptive Probing. In *Proceedings of the ACM SIGCOMM Conference*. ACM, Hong Kong, China, 255–266. https://doi.org/10.1145/2486001.2486017

[10] Lin Quan, John Heidemann, and Yuri Pradkin. 2014. When the Internet Sleeps: Correlating Diurnal Networks With External Factors. In *Proceedings of the ACM Internet Measurement Conference*. ACM, Vancouver, BC, Canada, 87–100. https://doi.org/10.1145/2663716.2663721

[11] Philipp Richter, Ramakrishna Padmanabhan, Neil Spring, Arthur Berger, and David Clark. 2018. Advancing the Art of Internet Edge Outage Detection. In *Proceedings of the ACM Internet Measurement Conference*. ACM, Boston, Massachusetts, USA. https://doi.org/10.1145/3278532.3278563

[12] Aaron Schulman and Neil Spring. 2011. Pingin' in the Rain. In *Proceedings of the ACM Internet Measurement Conference*. ACM, Berlin, Germany, 19–25. https://doi.org/10.1145/2068816.2068819

[13] Anant Shah, Romain Fontugne, Emile Aben, Cristel Pelsser, and Randy Bush. 2017. Disco: Fast, Good, and Cheap Outage Detection. In *Proceedings of the IEEE International Conference on Traffic Monitoring and Analysis*. Springer, Dublin, Ireland, 1–9. https://doi.org/10.23919/TMA.2017.8002902

[14] USC/ISI ANT project. 2017. https://ant.isi.edu/datasets/all.html. (2017). Accessed: 2019-01-08.