Evaluating Externally Visible Outages

USC/ISI Technical Report ISI-TR-701, Aug. 2015

Abdulla Alwabel[1] John Healy[2] John Heidemann[1]
Brian Luu[2] Yuri Pradkin[1] Rasoul Safavian[2]
1: USC/Information Sciences Institute 2: Federal Communications Commission

ABSTRACT

This technical report evaluates externally visible outages (those that can be seen from a third party) in the IPv4 Internet, comparing them to privately reported VOIP outages. Our primary goal is to understand if externally visible outages can complement current reporting mechanisms and provide a better understanding of the reliability of the U.S. telecommunications infrastructure. We measure outages with Trinocular outage detection system using active probing from multiple sites. Our secondary goals are to evaluate Trinocular for long-term use and potential operationalization, and to identify any problems and next steps needed in that direction.

1. INTRODUCTION

This paper evaluates externally visible outages (those that can be seen from a third party) in the IPv4 Internet, comparing them to privately reported VOIP outages. Our goal is to understand if externally visible outages can complement current reporting mechanisms and provide a better understanding of the reliability of the U.S. telecommunications infrastructure.

We draw upon the Trinocular outage detection system as a source of externally visible outages [1]. Trinocular uses ”pings” (ICMP echo request messages) from several sites to evaluate which public IPv4 networks are reachable. It has limitations—it cannot see behind firewalls, and it requires multiple addresses in each target network block to be willing to respond, but it has been previously evaluated and shown to be effective at detecting network outages. An earlier version has been used to study the effects of Hurricane Sandy on the New York/New Jersey area in 2012 [2].

In this report we compare Trinocular's public observations against network outages provided privately to the U.S. Federal Communications Commission (FCC) as part of their Network Outage Reporting System (NORS). NORS reporting is done by telecommunications providers when specific criteria are met (see Part 4 of FCC's rules, 47 C.F.R.). Because of the potentially proprietary nature of outage reporting, this data is confidential and not directly accessed outside of the FCC. Nevertheless, it serves as an important measure of U.S. telecommunications reliability.

The goal of this report is to compare these two sources information about network reliability. On the surface, they are quite different: one an open, external source of information about only the public, IPv4 Internet, the other private information provided by the network operators themselves, with a focus on voice and VOIP systems. Our goal is to understand how these sources can complement each other and lead to a better understanding of U.S. telecommunications reliability. Our secondary goals are to evaluate Trinocular for long-term use and potential operationalization, and to identify any problems and next steps needed in that direction.

2. DATA SOURCES AND CLEANING

We first compare our data sources and the specific period under study.

2.1 Comparison Period

We selected a 6-month period for study, from 1 April 2014 to 31 September 2014. We selected this period because it was recent at the time our study began and 6 months seems like a sufficient time to observe a range of activity.

2.2 About Public Trinocular Information

For Trinocular, this period corresponds to the datasets A16 [3] and A17 [4]. These datasets are available through the DHS PREDICT program (`https://www.predict.org/`) and from USC (`https://ant.isi.edu/datasets/`).

In this period Trinocular collected data from three locations: Marina del Rey, California (near Los Angeles), Ft. Collins, Colorado (near Denver), and Keio University, Japan (near Tokyo). We identify these sites as W, C, and J, respectively.

In the process of processing this data we worked around three data collection and analysis problems: timeouts in collection, occasional periods where there are two few vantage points active, and blocks that ”go dark” for extended periods (likely due to change in use or firewalls) and become unmeasurable. We define new data analysis methods that can correct each of these problems. The first two problems are under our control, and so from what we learn we identify operational changes that address the first two issues. (We have adopted these changes and confirmed the problems are addressed in subsequent datasets.)

2.3 Timeout Problems

Unfortunately, the version of Trinocular in use for most of this period had a latent bug that affected accuracy at the sites differently. The system strives to maintain a schedule where it probes every block every 11 minutes. When it gets behind, it would preemptively time-out pending queries. The result is that blocks were incorrectly declared down.

The degree to which incorrect decisions were made depends on the speed and loading of the measurement system.

Problems are worst at the site J, since it is older systems and it was modified to copy data off-site to archive it (to manage limited disk space).

Figure 1 shows the marginal distribution of how many blocks are down for each individual site for A16 (Figure 1a) and A17 (Figure 1b). In both datasets, site J is often unreliable with outages from 20-100% of blocks, but sites C and W are fine. This bug was fixed 2015-02-13 and we have confirmed that new datasets do not exhibit this problem.

2.4 Insufficient Vantage Points

Trinocular requires multiple vantage points to make strong conclusions that a block is down. For analysis here, we require at least three vantage points (VPs) to be active. (This requirement is a conservative choice, since any one vantage point can be affected by noise; requiring three makes it unlikely that noise at multiple points will occur at the same time.)

For A16 and A17 we were operating three vantage points. Thus, when one vantage point is unavailable we cease to report data. Figure 2a and Figure 2c show that we regularly have times where one or more VPs are down, as shown by the green peaks. These periods occur about every nine days, corresponding when our periodic automatic restart of Trinocular. We also see that there is one period of about a day on 2014-05-02 when one VP is down for an extended period. These times with insufficient VPs are regular but usually short (except for 2014-05-02). It is important to observe that this conservative measurement choice will result in bursts of ”new” outages, since a block that is down will transition from down to insufficient-to-measure back to down.

2.5 Unmeasurable Blocks

We have previously found that some blocks do not respond to our probes. Taken at face value, these blocks appear down. However, it seems unlikely that actively used blocks are actually down for weeks or months at a time. A more plausible explanation is that these blocks are filtering our probing traffic, or that they have been reassigned for different uses. We consider blocks that are down for more than 2 weeks to be unmeasurable, and we filter them out in two ways.

First, Trinocular uses Internet censuses with long-term probes to all addresses to form the list of blocks considered for outage detection [2]. This filtering is done before Trinocular runs.

Second, we filter Trinocular output after it runs. We identify blocks that are down 75% of the time or more for one week or more and remove them from analysis. This filtering focuses Trinocular on short-term outages. (Evaluation of the sensitivity of the results to these parameter choices is future work.)

We have implemented this filtering algorithm for the first time in this study (programs: outages_to_outagedownup and outagedownup_to_unmeasurable). Prior studies used shorter measurement periods (a few days, or a month), allowing analysis of the entire dataset to detect unmeasurability. However, the datasets A16 and A17 are each 3 months long, and we can see in the sampled plots of outages (Figure 3) that some blocks become unmeasurable (show up as a long-term outage) mid-way through the dataset.

To assess the impact of unmeasurable blocks, Figure 4 shows the CDF of what fraction of time is unmeasurable for all blocks that have any unmeasurable period. Each quarter has about 525k unmeasurable blocks, and about half of these are unmeasurable for nearly the entire period. Of the remaining, they seem to have no special duration of unmeasurability. From this we conclude two things. First, it is worth ignoring blocks that are almost always out, so we will remove blocks that are unmeasurable for more than 70 days (about 80% of the dataset). Second, there remain about 250k blocks that are unmeasurable for a significant amount of time. We will not factor the periods when they are unmeasurable into outages.

We can examine the marginal distribution in number of unmeasurable blocks over time in Figure 2, as the red line. We see that unmeasurability is fairly smooth, but has period steps approximately every 9 days. We believe these correspond with our automated periodic restarts of Trinocular. We also see that the trend over time is that the number of unmeasurable blocks rise somewhat. This result is expected; it will rise as we add networks to our opt-out list, or as networks choose to block our traffic. (Not shown on these graphs, but the number of unmeasurable blocks also rise over time: quarter we add in new blocks we have discovered through external Internet census taking.)

2.6 Trends in Outages: Marginal Distribution

Given the raw data of millions of blocks, we next need to find likely outages. We have done this with several different ways: (1) clustering the blocks in two dimension and plotting them [5] as conclusions, so we conservatively report no outages for that period.

We see several large peaks in the marginal distribution, suggesting large outages. For A16, we see a peak on 2014-05-20 and another on 2014-05-28. For A17, we see a very large peak on 2014-08-27 corresponding with the public Time Warner outage, and another on 2014-09-18. These peaks all stand out from daily noise.

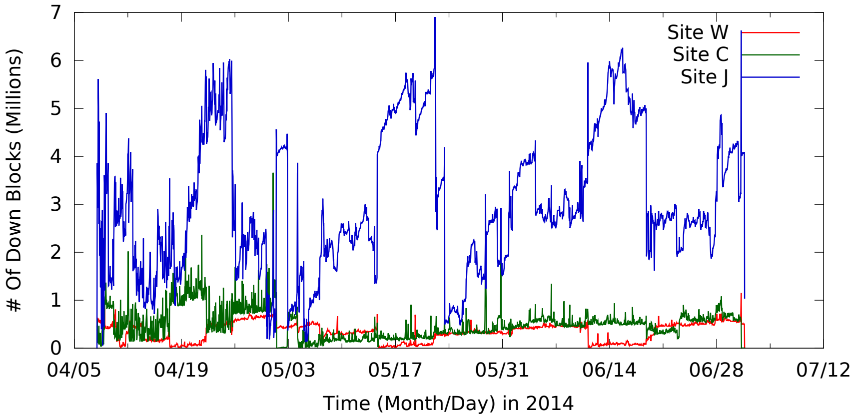
The first peak in A17 starts at Wed, 27 Aug 2014 09:00:00 GMT until 11:00:00 GMT of the same day, roughly two hours. This outage corresponds to Time Warner outages. The second peak in A17 starts at Thu, 18 Sep 2014 05:00:00 GMT until Thu, 18 Sep 2014 10:00:00 GMT and that corresponds to ’Uninet S.A. de C.V.,MX’ failure. For A16, the first peak starts at Mon, 19 May 2014 20:00:00 GMT and lasts for an hour. The second peak, starts at Wed, 28 May 2014 18:00:00 GMT and lasts till Wed, 28 May 2014 22:00:00 GMT.

2.7 Trinocular Restarts

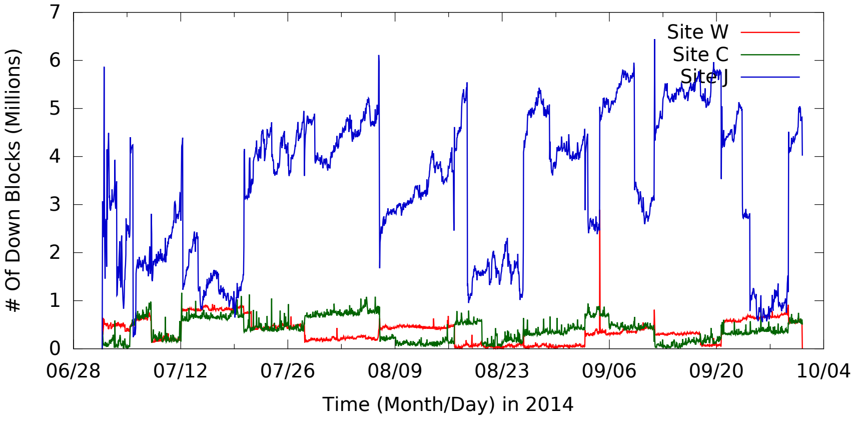
Trinocular is designed to run continuously and indefinitely (as long as its results are useful). As part of this process it automatically restarts itself at regular intervals. Originally restarts occurred every 5.5 hours; we extended that to just more than a week.

Unfortunately we discovered an interaction between these expected resets and long-term operation. We start a new series of outage datasets every quarter, and therefore start all sites at about the same time. With restarts happening at regular intervals after roughly the same start time, we all sites restart at about the same time.

Because we require at least three vantage points to be active to report outages (§ 2.4), this simultaneous restart

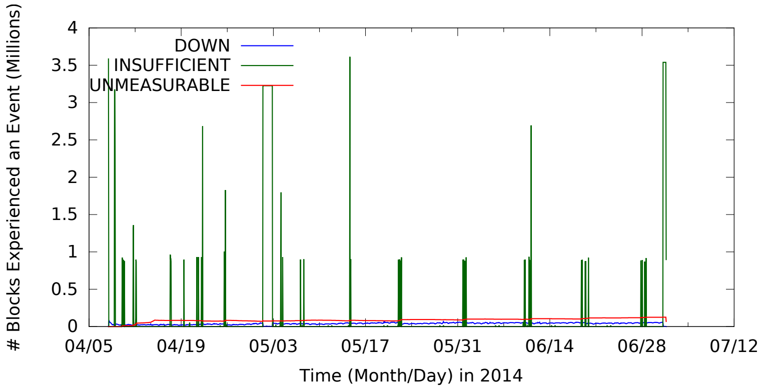


(a) A16

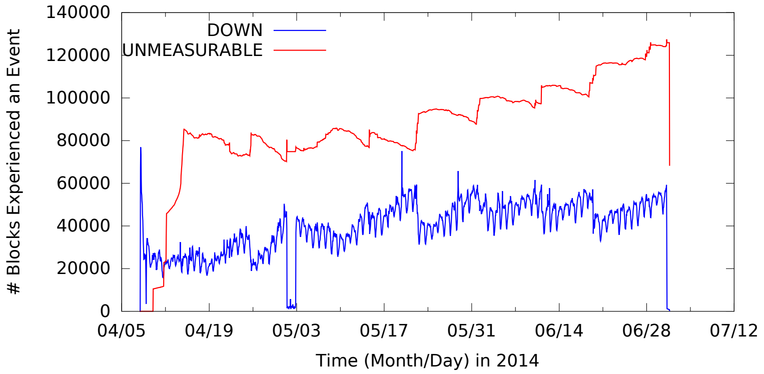


(b) A17

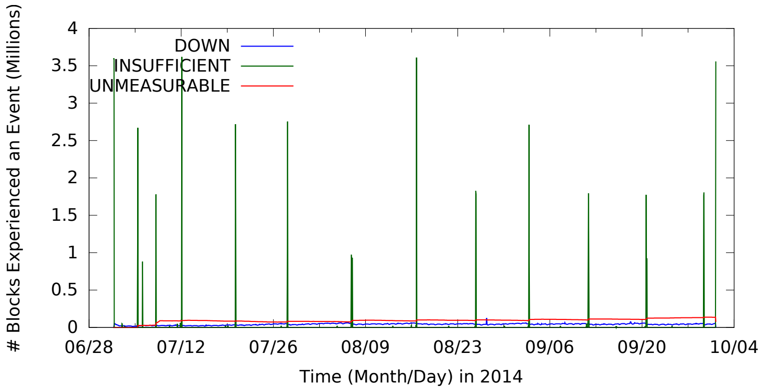
Figure 1: Marginal distributions of number of out blocks over time (Hours), per individual measurement site.



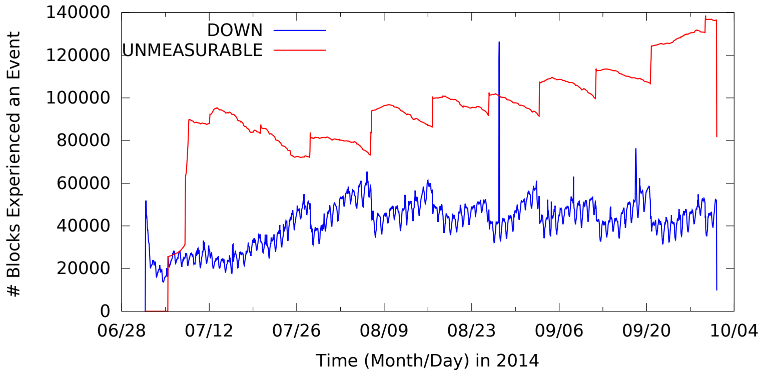
(a) A16; number of insufficient, down, and unmeasurable blocks.



(b) A16; number of down and unmeasurable blocks.

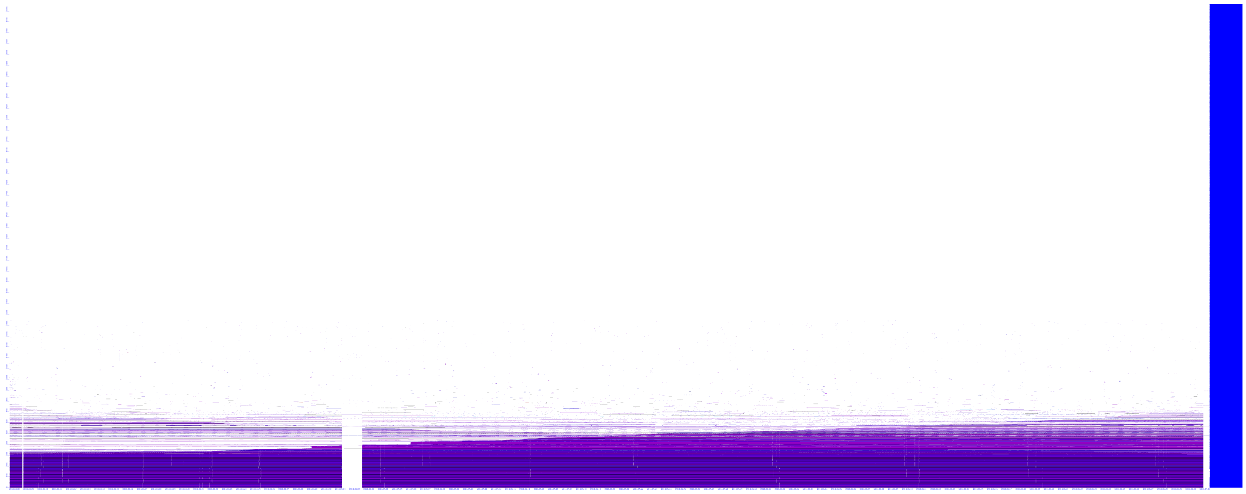


(c) A17; number of insufficient, down, and unmeasurable blocks.

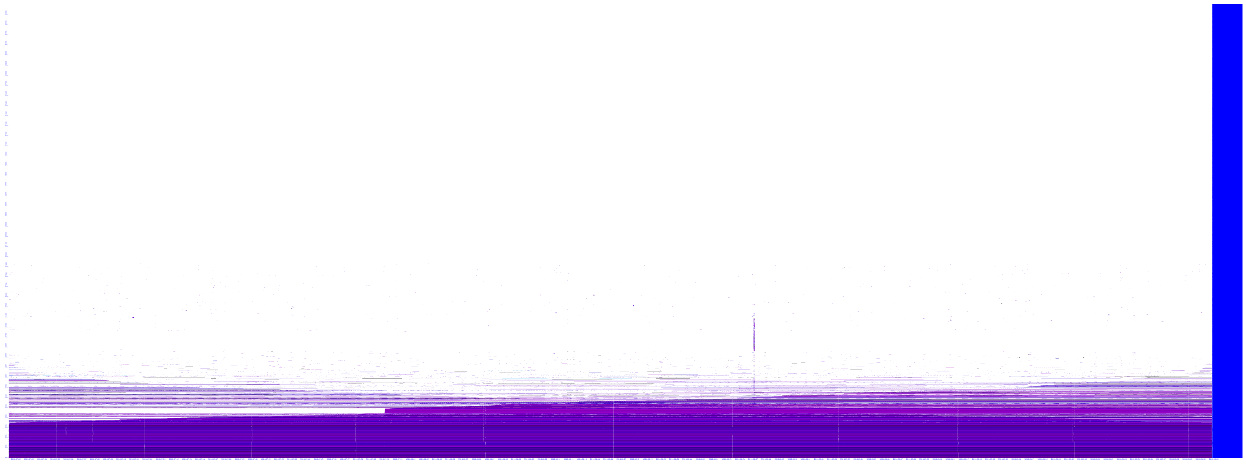


(d) A17; number of down and unmeasurable blocks.

Figure 2: Marginal distributions of number of blocks in different states over time. Datasets: A16 and A17.

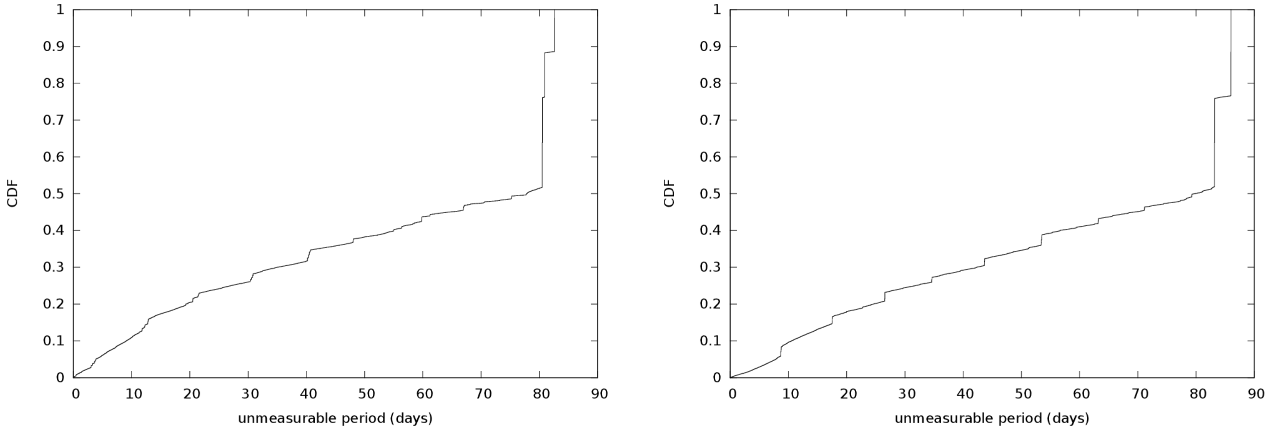


(a) A16



(b) A17

Figure 3: Visualization of 3 months of outages in A16 and A17. Both omit unmeasurable blocks, show U.S. blocks only, and are 1:200 sampling The vertical line about two-thirds of the way through A17 (bottom) is the Time Warner outage on 2014-08-26.



(a) A16; 530,342 blocks with some unmeasurability.

(b) A17; 596,786 blocks with some unmeasurability.

Figure 4: CDF of fraction of blocks where unmeasurability is less than a given number of days (only for blocks with some unmeasurable period.)

Zone	Longitude Range	Zone	Longitude Range
W11	[-172.5, -157.5)	E12	[172.5, 180),[-180,-172.5)
W10	[-157.5, -142.5)	E11	[157.5, 172.5)
W09	[-142.5, -127.5)	E10	[142.5, 157.5)
W08	[-127.5, -112.5)	E09	[127.5, 142.5)
W07	[-112.5, -97.5)	E08	[112.5, 127.5)
W06	[-97.5, -82.5)	E07	[97.5, 112.5)
W05	[-82.5, -67.5)	E06	[82.5, 97.5)
W04	[-67.5, -52.5)	E05	[67.5, 82.5)
W03	[-52.5, -37.5)	E04	[52.5, 67.5)
W02	[-37.5, -22.5)	E03	[37.5, 52.5)
W01	[-22.5, -7.5)	E02	[22.5, 37.5)
W00	[-7.5, 7.5)	E01	[7.5, 22.5)

Table 1: Longitude Ranges of Longitude Zones.

resulted in insufficient coverage for short times at the regular restart interval. This collection approach is visible in our data (see regular spikes in Figure 2) as spikes in the number of down blocks every 8.5 days (every 1110 cycles, each of 11 minutes).

We have corrected this problem by intentionally desynchronizing restarts. In the current version of our software each site runs for a slightly different duration before restarting. As a result, if all sites are initially started at the same time they are guaranteed not to restart at the same time. Over three months it is unlikely that even two sites will restart at same time, and if they do, they will do so once only and then be desynchronized. Preliminary analysis of dataset A19 [4] shows that these changes fix this problem.

3. ABOUT NORS INFORMATION

The Network Outage Reporting System (NORS) began collecting outage data in 2005. When NORS started, communication outages in wireline, wireless, cable telephony, paging, and satellite companies were reportable. In 2012, outage reporting was extended to companies providing VoIP services. Outages are reportable if they exceed specified thresholds. In general, outages have to last 30 minutes and affect 900,000 user-minutes (the product of the number of users affected and the duration of the outage in minutes). There were 15,818 final outage reports in 2014 of which 635 were VoIP outage reports.

4. PUBLIC TRINOCULAR OUTAGE INFORMATION

To help understand Internet outages, we visualize them by graphing the marginal distribution of the Trinocular data with unmeasurable blocks removed for each day (from UTC midnight to midnight), Figure 5.

For each day, we count outages grouped into 24 longitude zones, coloring each zone differently. A longitude zone corresponds to 15 degrees of longitude. They correspond to the solar time-of-day, but only approximate time zones which are influenced by political choices. We label zones as W11, W10,... W01, W00, E01, E02, ..., E11, and E12 with longitude ranges specified in Table 1. The W00 spans the longitudes of -7.5°to 7.5°. The W01 comprises of longitudes -7.5°to -22.5°. Similarly, the E01 encompasses the longitude range of 7.5°to 22.5°. Counts of outages in each zone are biased by variation in global population and Internet use (for example, there are few Internet users at longitude -180).

5. COMPARISON

To detect Internet outage event in the United States, we visualize the daily marginal distribution as described above but only for blocks belonging to US territories, see Figure 6.

To compare the NORS VoIP outage data to the Trinocular data, we determined the date in Greenwich time that each VoIP outage started. The number of VoIP outage reports ranged from zero to nine for each date. To best locate dates with unusual Internet outage, we first ignore all days having software restart since the restart of software using in Trinocular process causes significant false outages (which had been corrected for future collected data). Then we look for outages of data that are significant when comparing to the average of outages of dates in between software-restart dates.

We identify the following dates as having large Internet outages in the United States: 5/7/2014, 5/28/2014, 6/17/2014, 6/27/2014, 7/2/2014, 7/8/2014, 8/7/2014, 9/4/2014. We then consider two groups: these high-Trinocular-outage days and the other, low-Trinocular-outage days. Our goal is to understand if high-Trinocular-outage days correspond to days with large numbers of VoIP outages, suggesting that Trinocular can predict VoIP outages.

We tested whether the number of VoIP outage reports on high-Trinocular outage days was the same as the number of VoIP outage reports on low-Trinocular outage days using a t-test. The t-test rejected the hypothesis at less than the 0.001 level. This means we have very strong evidence that there are more VoIP outage reports on the high-Trinocular outage days than on the low Trinocular outage days. In other words when Trinocular data identifies large Internet outages, there is large chance that we will also see more VoIP outage reports.

In addition, there were a mean of 3.75 NORS VoIP outage reports on days with high-Trinocular outages, and only 1.74 on low-Trinocular-outage days. There is less than a 1 in a thousand chance that we would see this big a difference between the number of VoIP outage reports on high-Trinocular days and on low-Trinocular days purely by chance. There is a marked statistical difference in the number of VoIP outage reports on these two types of days. Days identified by Trinocular with a large number of Internet outages have more VoIP outage reports than days with a low number of Internet outages.

6. DISCUSSION AND NEXT STEPS

Goals of this pilot project are to evaluate the use of Trinocular to detect Internet outages and to better understand the relationship between Internet outages detected by Trinocular and the FCC's NORS system. This work has been successful in both these goals, with several important findings.

We have *improved the robustness of Trinocular* in several ways. The transition from a research prototype to a robust tool requires refinement and scrutiny. This 6-month study helped us *adapt tools to scale to processing months of data at a time*, from their prior use for relatively short datasets (a few days or weeks). An important transition was to identify unmeasurable blocks and to track blocks that become unmeasurable over the course of observation. While one can treat the population of a week-long dataset as stationary, without blocks being added or removed, this assumption must be relaxed when we examine months of data.

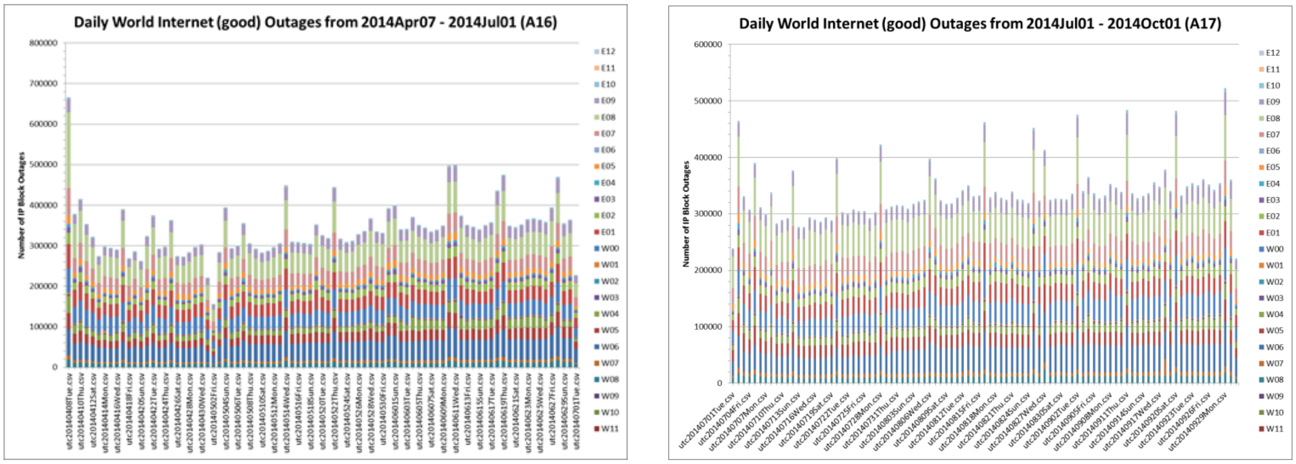


Figure 5: Daily Internet outages using Trinocular data without unmeasurable blocks.

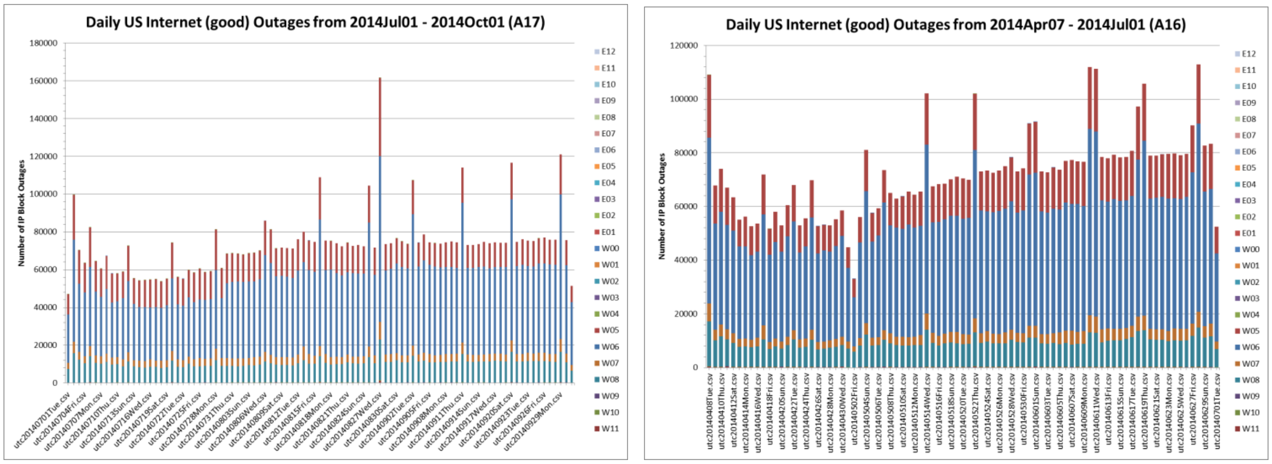


Figure 6: Daily US Internet Outages based on Trinocular Data without Unmeasurable Blocks.

We also identified and fixed several problems in Trinocular: a bug that caused data loss under high load and correlations in restart times that resulted in brief periods of unmeasurability.

More importantly, we *have established the relationship between Trinocular outages and FCC reports in NORS*. We found statistically strong relationships between days of large numbers of Trinocular-observed outages in the Internet with reports of VoIP outages in NORS. This result documents the ability of external, third-party measurements to estimate the reliability of the U.S. telecommunications infrastructure, at least qualitatively.

Overall, these results suggest that *Trinocular can be useful as a new tool for the FCC*. As use of Internet-based telecommunications grows in the U.S. telecommunications industry, this ability to observe network reliability will be of increasing importance.

Next Steps: Although this study accomplished its goals, it also identified several potential directions for additional study:

Studies at finer timescales: We considered data only a daily basis. It may be possible to examine outages with more precise timing. The limits here are the precision of Trinocular data, with at most 11 minute precision in standard use, and the precision of NORS reporting data (reported in hours, but sometimes with uncertainty about timezones). We believe analysis of 4- to 8-hour blocks of time may be feasible.

Improved geolocation and outage event identification: Additional work could be done to determine where outages occur and show correlations in outage location. The fundamental limit here is the quality of Internet geolocation data.

In addition, automatic identification of outage events is important to find individual outage events. This report studied marginal distributions and only identified very large outages in this aggregate data, but we have successfully identified outage events in smaller datasets. We would like to adapt our event discovery tools to full-size, long-term data.

Studies with current Trinocular: Our studies here resulted in several improvements to Trinocular. Newer Trinocular observations that include these improvements may allow a cleaner comparison.

Additional improvements to Trinocular: Success of this early comparison suggests there may be benefit in improving Trinocular to report results more quickly.

Visualization of outages: Our current approaches to visualizing outages are relatively limited. Detection of outage events and near-real time outage reporting would allow outages to appear on a world map, nearly as they occur.

7. CONCLUSIONS

This study demonstrated that Trinocular can be used to track Internet outages for months at a time, and that Trinocular-observed outages in the public Internet are statistically related to VoIP outages are reported in the FCC NORS system. This work suggests it would be beneficial to continue to develop Trinocular to augment existing FCC tools to report network problems and to help oversee the U.S. telecommunications system.

8. REFERENCES

[1] J. Heidemann, L. Quan, and Y. Pradkin, "A preliminary analysis of network outages during hurricane sandy," Tech. Rep. ISI-TR-2008-685b, USC/Information Sciences Institute, November 2012. (correction Feb. 2013).

[2] L. Quan, J. Heidemann, and Y. Pradkin, "Trinocular: Understanding internet reliability through adaptive probing," in *Proceedings of the ACM SIGCOMM Conference*, (Hong Kong, China), p. to appear, ACM, August 2013.

[3] USC/LANDER project, "Internet address outage dataset, predict id `usc-lander/internet_outage_adaptive_a16all-20140407`." web page `http://www.isi.edu/ant/lander`, 2014.

[4] USC/LANDER project, "Internet address outage dataset, predict id `usc-lander/internet_outage_adaptive_a17all-20140701`." web page `http://www.isi.edu/ant/lander`, Oct. 2014.

[5] L. Quan, J. Heidemann, and Y. Pradkin, "Visualizing sparse internet events: Network outages and route changes," in *Proceedings of the First ACM Workshop on Internet Visualization*, (Boston, Mass., USA), p. to appear, Springer, November 2012.