Towards Improving Outage Detection with Multiple Probing Protocols

Manasvini Sethuraman, Zachary S. Bischof, and Alberto Dainotti

Georgia Institute of Technology

Abstract. Multiple systems actively monitor the IPv4 address space for outages, often through ICMP probing. In this work, we explore the potential benefits (in terms of increased coverage) of leveraging additional protocols for probing by analyzing Internet-wide scans conducted using transport layer (TCP/UDP) probes. Using several existing Internet-wide scan snapshots, we show that between 531k to 606k additional /24 blocks, which were originally too sparse to be monitored via ICMP probing alone, now have the potential to be monitored for outages. We also find that it is possible to improve the probing efficiency for 850k-970k blocks, of which, 106k-125k blocks were not observed in the previous two years of ICMP-based scans. We observe that the average percent of /24 blocks per AS that could potentially be reliably monitored for outages increases from 65% to 83%, spanning 28k ASes.

1 Introduction

The Internet is critical communications infrastructure. As such, the ability to monitor connectivity and detect outages is of exceptional importance. An effective approach to detecting Internet outages requires broad coverage (i.e., not limited to a particular ISP or protocol), accuracy (i.e., few false positives and false negatives) and speed (i.e., rapid identification of outages). Active probing techniques have been demonstrated to be particularly successful in meeting some of these requirements and are currently deployed in various operational outage detection systems [18, 2]. However their coverage of the address space is in some cases limited, and—as we show in this study—this is partially due to their exclusive use of ICMP probes.

This paper studies the potential for improving active measurement based outage detection coverage by supplementing ICMP measurements with transport layer (TCP, UDP) probing. In the context of network discovery and scanning, previous work has shown that leveraging multiple protocols (e.g., ICMP, UDP, TCP) allows for discovering the largest number of active addresses in use [6]. In this paper, we explore if and to what extent this property applies in the context of outage detection, where certain prerequisites—such as stability over time of address block responsiveness—are fundamental.

To estimate the potential benefits of a multi-protocol approach in practice, we use Trinocular as a reference outage detection method [18, 5]. Trinocular is a well-studied active probing methodology that uses ICMP probes to detect outages at

the /24 block granularity using Bayesian inference. The methodology is currently deployed in operational outage detection systems, including IODA [2].

In order to determine which address blocks can be monitored for outages with active probes, Trinocular analyzes historical full Internet scan data over the preceding three years to identify responsive hosts and calculate the average availability for each /24 block. To simulate augmenting the calculation of /24 block availability to include the results of transport layer measurements, our analysis incorporates two existing datasets: ICMP probing results from ISI ANT's Internet history dataset [13] and snapshots of Internet-wide transport layer (TCP/UDP) scans from Censys [8]. By leveraging both of these datasets, we are able to quantify temporal variations in host responsiveness for each protocol individually as well as the resulting combined set of hosts. Analyzing two years of data on host responsiveness from both datasets, we calculate the potential improvements for outage detection coverage in terms of /24 block discovery and host availability that can be obtained by supplementing ICMP probes with transport layer probing.

The contributions of our analysis[1] can be summarized as follows:

- Hosts that respond to a combination of transport layer and ICMP probes have, on average, 10% higher availability than hosts that respond only to ICMP probes.
- Incorporating transport layer probing can improve coverage in various ways: (i) 845k blocks defined "unreliable" in Trinocular (i.e., with availability <30%) experience an increase in availability from an average of 13% to 48%; (ii) an additional 320k blocks that were not seen at all via ICMP, of which, between 106k-125k are "reliable". As a baseline for comparison, note that with ICMP only, Trinocular finds 3.5M reliable blocks.
- We find that such a non-negligible improvement – i.e., a relatively small percentage but over a very large set (i.e., the Internet) – is particularly pronounced in certain ASes, thus illuminating portions of the Internet where outage visibility would otherwise be limited. E.g., the blocks discovered only through transport layer scans span 14k ASes, of which 1.8k ASes are newly discovered through the additional probes.
- Improvements in outage detection coverage are exceptionally notable in certain geographic regions, often in the Global South. For example, Ethiopia, Suriname and French Polynesia all have a relatively small number of blocks that can be reliably monitored via ICMP-only probing (26, 25 and 37 blocks respectively). In each case, the addition of transport layer probes increases the number of blocks that can be reliably monitored to over 200 blocks.

2 Related Work

Detecting network outages via active measurements has received significant focus from the networking research community. Prior work, such as Fan and Heidemann, studied the use IPv4 hit lists to make scanning more efficient for surveying

[1] Repository: https://github.com/InetIntel/ioda-censys-isi

the topology of the Internet [10]. The authors demonstrate that selecting appropriate representative hosts can improve traceroute-based coverage of edge hosts. Later efforts used this idea, such as in the Trinocular methodology [18], which proposed techniques for outage detection probing hitlists with active ICMP measurements in 11-minute cycles. However, /24 blocks that contained a small number of hosts (i.e., blocks with less than 15 hosts responding to probes) were difficult to accurately monitor. In subsequent work that aimed to reduce the impact of sparse blocks in identifying outages, Baltra and Heidemann proposed full-block scanning (FBS) as an extension to Trinocular [5]. However, FBS involves probing every address in the /24 block and may take several probing cycles to complete.

In addition to Trinocular, there are a number of efforts that aim to detect Internet outages. Long-running TCP probes within the RIPE infrastructure [4] are used to infer outages in DISCO [22]. Thunderping examines outages due to adverse weather conditions in residential networks by probing against a sample of hosts in a geographical area, before and after adverse weather [21, 17].

Passive measurement techniques using various data sources for inferring outages have been proposed in literature—including Internet Background Radiation [11], NTP server events [23], and HTTP logs [19]. However, they are more commonly used in conjunction with other approaches. IODA uses several signals (active ICMP probes, passive monitoring of BGP prefix announcements, and passive monitoring of Internet background radiation) to detect outages across the world [2]. Hubble uses a combination of active measurements (ICMP probing) and passive monitoring (BGP prefixes) to infer connectivity problems on the Internet [14].

With regards to enumerating hosts and identifying liveness in the address space, Bano et al. use a combination of TCP, UDP, and ICMP probes to create a point-in-time snapshot of the Internet and analyze host activity that is visible via different protocols [6]. They investigate the liveness of hosts, under different protocol probes and the correlation in responses between different protocols. However, they do not investigate the implications of TCP/UDP probes on a block's availability and do not analyze the longitudinal trends in availability of a host or a block. On a similar vein, IRLScanner investigates host liveness and service discovery via Internet-wide scans comprised of different protocols (DNS, HTTP, SMTP, EPMAP, ICMP, and UDP ECHO) [15].

3 Datasets and Methodology

In this section, we first briefly describe the datasets that we use in our analysis. We then provide a high level summary of our findings (Section 3.1) and discuss metrics for comparing the coverage of the datasets (Section 3.2).

Our analysis in this paper primarily relies on data from two sources: IPv4 censuses collected by the ANT Lab at ISI [1] and IPv4 scans conducted by Censys [8] using Zmap [9]. For both datasets, we use scans collected between November 2020 and December 2022. We do not consider specific vantage points which elicited

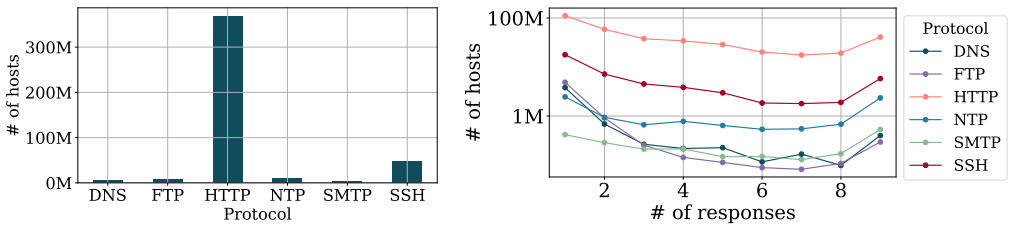
a response, since we are primarily interested in whether or not measurements elicit a response from a particular IP, rather than the route to a specific host. Further, while IP address churn is possible, we are interested in knowing which hosts consistently respond to probes across multiple surveys, and are likely to respond in future measurement cycles, and may be informative for detecting outages.

ANT dataset (ANT): As part of the ANT Censuses of the Internet Address Space project [12], researchers at ISI conduct periodic surveys of the IPv4 address space, recording responses from each host. Hosts are probed using ICMP and the response is recorded as a 0 or a 1. A zero could indicate either error, not probed or not responded, and a 1 represents an echo response. Historical data of host responses is made available through the IP history dataset. This dataset contains responses from every address from the previous 16 surveys, starting in 2011. Surveys are typically about 2-3 months apart with each survey taking several weeks to complete [12].

Censys Internet-wide scan data (Censys): Censys provides researchers access to its Internet scan data. Snapshots in this dataset contain Internet-wide scan data at roughly one week intervals. The scanning method is based on ZMap [9], with each scan taking about 45 minutes to complete. Entries in a snapshot contain information about the responding IP address, the services running on the host, and other information (exposed ports etc.) that are identified using a variety of handshakes such as TLS, HTTP, SSH and TELNET. A host could respond to several protocol handshakes, and we see this quite often in the snapshots. We record the list of protocols for which a response was received to aid us in constructing our dataset.

Censys scans contain the results from more than 30 protocol handshakes. For designing a probing system that is fast, making multiple protocol handshakes is time consuming. Instead we consider those protocols which produce responses from the maximum number of hosts. We narrowed down this list of protocols to HTTP, SSH, FTP, NTP, SMTP and DNS. For each host, we then find the protocol which yields the most number of responses across the nine scans. The full set of protocols in Censys produces responses from 467M across nine scans. This number drops to 440M hosts, or 94% of the original set of Censys hosts when we limit the number of probing protocols to one per host. From Fig. 1(a), we see that HTTP probes elicit responses from well over 300M hosts, followed by SSH with 45M hosts, and the remaining protocols with fewer than 10M hosts, across all nine snapshots. In Fig. 1(b) we see that there is a decrease in the number of hosts responding to two or more surveys up until 8 surveys, and then an increase in the number of hosts responding across all nine surveys.

For our analysis, using the ANT and Censys scans to which we had access at the time we conducted our analysis, we match each ANT survey to a Censys scan such that the difference between the two start dates is as small as possible. We use nine matched ANT and Censys scans between November 2020 and December 2022. While there are sometimes differences between the start dates of the corresponding ANT and Censys snapshots, these differences are relatively small



(a) **(b)**

Fig. 1: (a) Distribution of best protocol responses across hosts across all nine snapshots (b) Variation of number of hosts responding to a protocol handshake with number of surveys.

in comparison to the long time frames of the ANT censuses. Additional details on these scans are available in Appendix A.

3.1 High Level Observations

We first investigate how many additional blocks can be discovered via Censys' probing methodology and if existing (and new) blocks can be reliably probed using other protocols by collating the ANT and Censys datasets. On average, over nine surveys, we find that ANT covers 660M hosts across 5.8M /24 blocks, Censys covers 467M hosts spanning 5.78M /24 blocks, and the combination of the two datasets covers 820M hosts spanning 6.21M /24 blocks, if we consider probing every host using 30 different protocols. We will refer to this dataset as "Censys-any". If we only consider the best protocol for each individual host, which we call "Censys-best", we get 440M hosts, spanning 5.72M /24 blocks. When we consider ANT and Censys-best together, we get 810M hosts spanning 6.21M /24 blocks.

Over time, we noticed that the number of ASes covered by ANT went from roughly 64k in November 2020 to over 69k in December 2022. The combination of the two datasets yielded 67k ASes in November 2020 and 71k ASes in December 2022. More information on these summary statistics can be found in Appendix A. Appendix A.

3.2 Key Terms and Metrics for comparison

In our analysis, we compare the coverage of ANT and of the combined "ANT + Censys" datasets using metrics such as block availability and sparsity, as defined in the Trinocular [18] work. We will add -any or -best to Censys to indicate which method was used to elicit responses from hosts. In this section, we first define these and other related terms.

Availability. In the outage-centric model of the Internet proposed in the Trinocular, historical block availability is used to seed the outage detection model. Block availability, or $A(E(b))$ is defined as the response rate of the block averaged over the last four surveys, for all the hosts that ever responded in that block in the last three years. This metric is used as an initial estimate to seed the calculation of a block's status (up/down). $E(b)$ represents the number of

hosts in a block that ever responded to any probe. As a simplified example, consider two probing cycles. In the first cycle, only one host from the block responded, and in the next cycle, a different set with nine hosts responded. In this case, over the two cycles, $E(b)$ would be 10. $A(E(b))$ for the block would then be 0.5, since all of the hosts that were ever observed in any cycle responded in one of the two probing cycles. Since we only had access to approximately two years of Censys snapshot data (from November 2020 to December 2022), we modify the definition of $A(E(b))$ to only consider hosts which responded to any probe over the approximately two year period for both datasets. We used all the snapshots we collected for computing availability scores. Trinocular computes $A(E(b))$ from ICMP probes in ANT censuses (lasting several weeks) to generate initial probabilities for block statuses. We extend this methodology to incorporate TCP/UDP probes from Censys in addition to ICMP probes.

Sparse block. A block that has fewer than 15 responding hosts in a particular survey is labeled as a sparse block. Blocks with 15 or more responding hosts are referred to as nonsparse blocks.

Reliable block. Among non-sparse blocks, there are blocks that require fewer rounds of probing to infer their status. Specifically, when a block's availability $A(E(b))$ is at least 0.3, Trinocular is able to *always* detect an outage which lasts longer than one round of probing ($\approx$11 minutes). We refer to the blocks with $A(E(b)) \geq 0.3$ as reliable blocks. Blocks with $A(E(b)) < 0.3$ are referred to as unreliable blocks, and it may take several rounds of probing to establish the status of such blocks.

Upper/lower bounds on sparsity. ANT surveys and Censys scans do not typically start on precisely the same day and take different periods of time to complete. This makes a direct comparison of sparse blocks difficult, as differences could be due to DHCP, hosts being restarted, etc. In order to avoid over-estimating the reduction in the number of sparse blocks by adding Censys probing techniques, we calculate a lower and upper bounds for the number of sparse /24 blocks when combining the two data sources.

For each matched survey, we consider the combined set of hosts in each block from both ANT and Censys. We then compute the number of sparse blocks from the combined data to get an optimistic lower bound on the number of sparse blocks. To calculate an upper bound, we compute the list of sparse blocks from the ANT dataset and then locate these blocks in the Censys data, checking for sparsity in the Censys dataset. If the blocks would be considered nonsparse using only the Censys data, we exclude these blocks from the list of sparse blocks. The final list of sparse blocks in either scenario is expected to be smaller than the original list of sparse blocks in ANT.

4 Impact at the Host Level

We first investigate if incorporating Censys' probing techniques could help identify hosts that can be reliably probed. Such hosts can be used when building hit lists, potentially allowing for easier and more rapid detection of outages.

Metric	IPs unique to ANT		Common IPs/ ANT		Common IPs/ ANT + Censys	
	Censys-Any	Censys-Best	Censys-Any	Censys-Best	Censys-Any	Censys-Best
Avg. response count	5.29	5.24	5.20	5.25	6.12	6.02
Prob. of low response count	0.17	0.17	0.15	0.15	0.03	0.04
Prob. of high response count	0.25	0.24	0.22	0.22	0.27	0.28

Table 1: Average number of responses for hosts across 9 surveys. Combining datasets increases the number of responses.

We calculate the number of responses of each host across 9 surveys as the number of times it responded to a probe, meaning each host has a response count between 1 and 9. We summarize host responses for different views of the data: *(i)* the average number of responses using ANT data for IP addresses that were unique to ANT (i.e., did not appear in any Censys scan); *(ii)* the average number of responses using only ANT data for IP addresses that appeared in both the ANT and Censys scans (though not necessarily in the same matched survey); and *iii)* the combined average number of responses for the IPs common to both datasets, with each IP considered responsive for a survey if the IP appeared in a corresponding ANT or Censys scan. For each category, we also calculated the fraction of hosts with a low response count (i.e., only appeared in 1 of the 9 surveys) and high response rate (i.e., appears in all 9 surveys). We consider two cases for Censys: using the dataset constructed from using any protocol to probe an address (Censys-any), and using only one protocol per address that yielded the most number of responses across the surveys (Censys-best).

Table 1 summarizes our findings related to host responses. On average, a host that is present in either dataset provides a response in 6 surveys (for both the Censys-Any and Censys-Best subsets), compared to 5.2-5.3 surveys for a host when only considering the ANT dataset. More importantly, the fraction of hosts that appear in both datasets with a low response rate, decreases from 15% to 3-4% when Censys data is added. Finally, adding Censys data also increases the fraction of highly available hosts from 22% to 27-28% for the hosts that are common to both datasets.

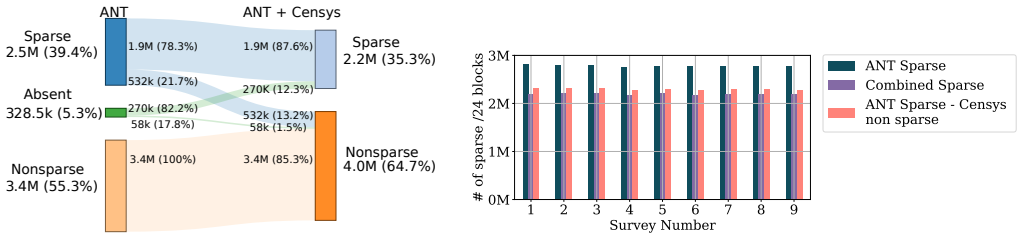
Takeaway Hosts present in both datasets are likely to appear in more surveys than hosts present in only one of the datasets.

5 Impact at the /24 Block Level

In this section, we study how adding Censys data can impact sparsity and reliability at the /24 block level.

5.1 Sparsity

We classify blocks as sparse (or nonsparse) based on the number of hosts found in the block (as defined in Section 3). For blocks that appear in the Censys



(a) **(b)**

Fig. 2: (a) In the most recent snapshot, adding hosts from Censys reduces the number of sparse blocks. (b) ANT sparse (blue bars) shows the number of sparse blocks in the ANT dataset in each survey. Combined sparse (purple bars) shows the lower bound on number of sparse blocks in the combined dataset (Censys-best). The pink bars show the upper bound on the number of sparse blocks in the combined dataset (Censys-best).

dataset but not in the ANT dataset, we denote the block as absent. Using the most recent snapshot of the data, we calculate the number of sparse, nonsparse, and absent blocks for both the ANT and combined (ANT and Censys) datasets. Fig. 2a shows that 58k (18%) of the 328k blocks that are completely missing in the ANT dataset become nonsparse by adding Censys-best. Further, 532k (21%) blocks which were sparse in the ANT dataset are no longer sparse after adding Censys-best.

For each snapshot, we compute the upper and lower bounds (defined in Section 3) for the number of sparse blocks after the addition of Censys-best data (Fig. 2b). We observe that sparsity decreases somewhere between 17% and 21%. The upper and lower bounds on the reduction in the number of sparse blocks are within 4% of the total number of sparse blocks originally found in the ANT data.

Takeaway Combining the two datasets yields between 488k to 600k fewer sparse blocks in comparison to the ANT dataset, or 17-21% reduction in the number of sparse blocks. Using Censys probing techniques has potential to increase the number of blocks that can be monitored for outage detection and may improve the accuracy and rapidness for other blocks.

5.2 Reliability

We consider the universe of hosts H to be the set of hosts discovered through either probing method. We then compute $A(E(b))$ using only ANT and again using the combined dataset, for each /24 block, where the elements of b are members of H. We classify the blocks as reliable or unreliable based on $A(E(b))$. We denote blocks with zero availability, or blocks not found in the ANT dataset as absent. Fig. 3 shows the Sankey diagram before and after combining the two datasets. It shows that 720k, or 31% of the blocks which were unreliable in ANT (2.3M) can now be efficiently probed for outages after adding hosts from Censys-best. Another 106k (32%) of 328.5k blocks which were missing in ANT, now can be probed efficiently for outages in the combined dataset.

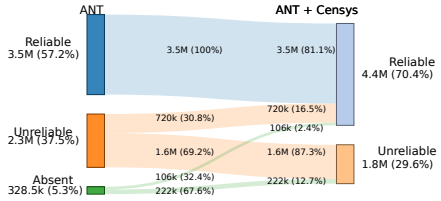
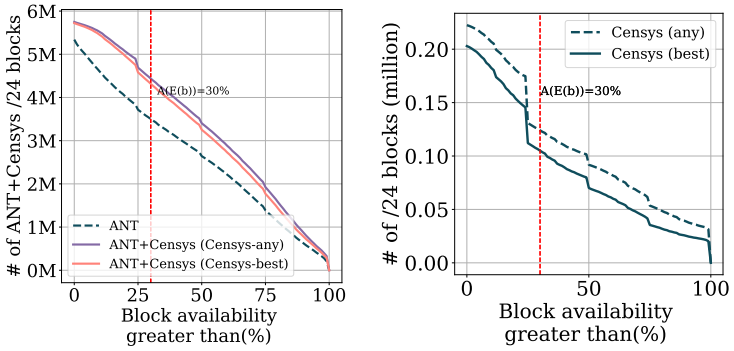


Fig. 3: Adding hosts from Censys improves representation and block availability.



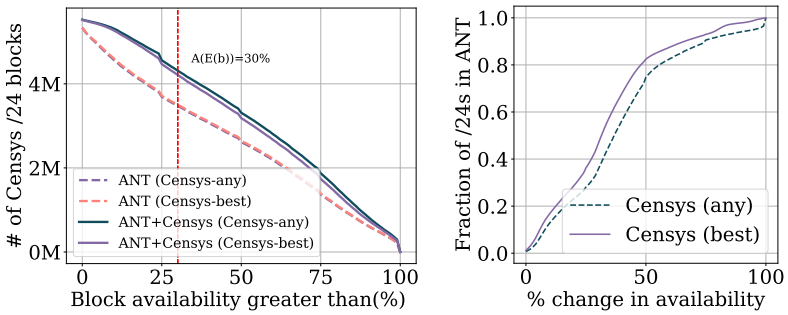
(a) **(b)**

Fig. 4: (a) Number of /24 blocks in either dataset with an availability score above a particular percentage for the ANT and combined datasets. (b) Block availability for /24 blocks found only in the Censys data. The vertical red dashed line in each plot represents the $(A(E(b)) = 0.3)$ **threshold for reliable outage detection.**

Improvement in reliability of existing blocks. Fig. 4a illustrates the improvement in block availability as a consequence of adding Censys-any/Censys-best data for all blocks (appeared in either dataset). Using only ICMP probing, 56.5% (3.5M) of all (6.2M) /24 blocks are reliable. Incorporating Censys data, increases the percent of reliable blocks to 70% using Censys-best (4.37M blocks) to 72% in Censys-any (4.48M blocks). Fig. 5a illustrates the block availability for only the blocks that appeared in the ANT dataset (5.8M blocks). As expected, the gains from Censys-any are slightly higher than Censys-best, since we use only one additional probe per host in Censys-best.

For the blocks which originally were not reliable in ANT, and are reliable in the combined dataset, we compute the change in availability. On average, we see an increase in availability from 13% to 50% across 845k /24 blocks, with Censys-any. With Censys-best, we get an increase in availability from 13% to 49% across 720k /24 blocks. Fig. 5b shows that about 25% of /24 blocks see at least a 50% increase in $(A(E(b))$.

Discovery of New Blocks. Adding Censys data also increases the number of blocks that we discover. After computing $A(E(b))$ for the newly discovered blocks, we find that 106k (Censys-best) to 125k (Censys-any) can be reliably



(a) **(b)**

Fig. 5: (a) Block availability for /24 blocks discovered in ANT (b) Change in availability of previously unreliable blocks in the ANT ISI dataset upon addition of hosts from Censys scans.

probed for outages as shown in Fig. 4b. These blocks were previously unresponsive and make up 2-3% of the final set of all reliable blocks.

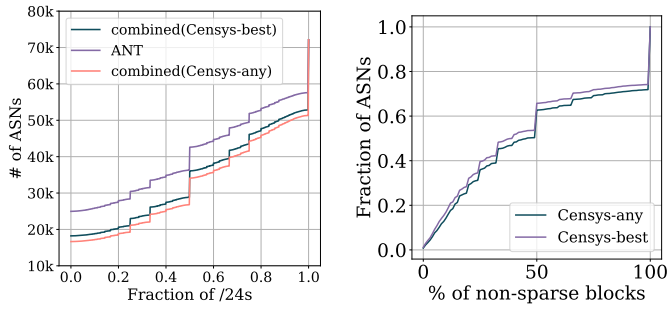
Takeaway. On average, $A(E(b))$ for a block b increases from 0.45 to between 0.51 (Censys-best) and 0.54 (Censys-any) when we consider host probing when we use a combination of probing methods. Furthermore, an additional 857k to 974k /24 blocks can now be reliably monitored for outages in a single Trinocular probing cycle and do not require full scanning (e.g., via FBS). Finally, 100-125k previously unseen blocks can now be probed reliably.

6 Impact at the AS Level

In this section, we explore the impact of combining the ANT and Censys datasets at the level of autonomous systems. We look up the AS associated with a /24 through CAIDA's pyipmeta [7]. We find 71k ASes, from both data sources combined, making up 76% of the 94,935 assigned AS numbers from NRO's extended delegation file published on April 14, 2023 [16].

6.1 AS-level reduction in sparsity

For the most recent snapshot of ANT and combined datasets, we group /24 blocks by their origin AS. We then find the number of nonsparse blocks per AS for the ANT dataset. For the combined dataset, we find the upper bound of the number of sparse blocks per AS and subtract these from the total number of blocks associated with the AS. Fig. 6a shows the number of ASes that have a particular percentage of nonsparse blocks for the ANT and combined dataset. We find that 60%(Censys-best) to 62% (Censys-any) of ASNs have at least 50% nonsparse blocks, in comparison to 50% of ASNs when using only the ANT dataset. Fig. 6b shows the CDF of the percentage of nonsparse blocks grouped by AS for the blocks unique to Censys. On average, 48% (Censys-best) to 51% (Censys-any) of the blocks per AS are nonsparse.



(a) (b)

Fig. 6: (a) CDF of % of non sparse /24 blocks per AS for blocks originally in ANT versus blocks that are non sparse in at least one of the datasets (ANT or Censys) for two cases: when using any protocol in Censys (pink line) and when using only the best protocol for the fiven host (blue line) (b) CDF of % of non sparse /24 blocks per AS for blocks discovered uniquely through Censys

Takeaway. On average, 53%(Censys-best) to 55% (Censys-any) of the blocks belonging to an AS are nonsparse in the combined dataset, in comparison to 44% of blocks per AS in the ANT dataset in the most recent snapshot of the datasets.

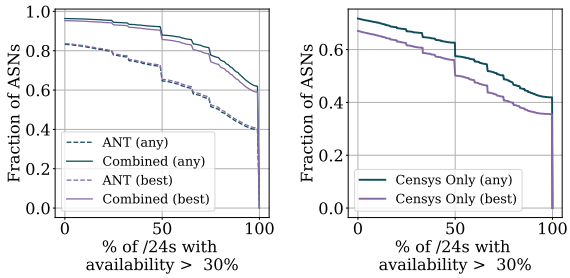
6.2 Availability at the AS level

We consider the effect of increased availability at the AS level on blocka already covered by ANT, and new blocks discovered by Censys.

Existing blocks. We compare the number of reliable blocks within an AS, when only ICMP probing is used and when a combination of probing methods is considered. ASes vary widely in size, so we plot the CDF of relative increase in the number of reliable /24 blocks per AS, for those ASes already discovered by ANT (Fig. 7a). We find that within an AS, 83% (Censys-best) to 86% (Censys-any) of /24s are reliable on average, when we consider both ICMP probing and internet scanning, in comparison to 66% of reliable /24 blocks per AS if we only consider ICMP probing. Out of the 12,559 ASes which did not have any reliable /24 block originally in the ANT dataset, 8,908 (Censys-best) to 9,732 (Censys-any) of the ASes have at least one reliable block when Censys hosts are added.

New blocks. The CDF of the number of reliable /24 blocks per ASN from Censys is shown in Fig. 7b. 9,849 to 10,404 ASes are covered by 0.106M to 0.125M reliable blocks unique to Censys-best and Censys-any respectively. Of these, 1,854 ASes were unique to Censys-any, covering 4,767 reliable /24 blocks, and 1,838 ASes were unique to Censys-best, covering 4,647 reliable /24 blocks.

Takeaway. Combining the two datasets results in an increase in the number of reliable blocks per AS from 65% to 83-86% on average.



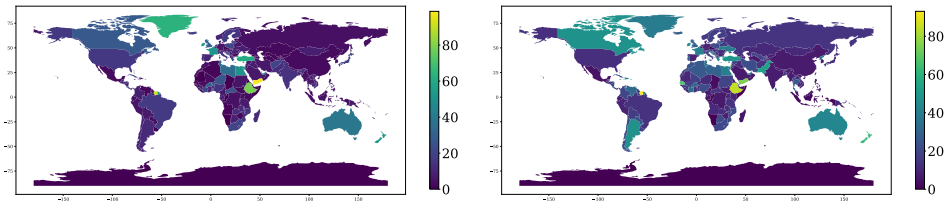
(a) (b)

Fig. 7: (a) CDF of % reliable /24 blocks per AS for all ASes discovered through combination of ANT ISI and combination of ICMP and other protocols for the same blocks (b) CDF of % of reliable /24 blocks per AS for all blocks discovered uniquely through Censys.

7 Block Geolocation

We investigate which countries and continents experience reduction in sparsity and improvements to reliability. We used the commercial version of the IpInfo [3] data to geolocate /24 blocks found in ANT and in the combined dataset.

Geolocating sparse blocks. For the most recent snapshot of ANT and the combined (Censys-best) dataset, we compute the number of sparse blocks by country. For the combined dataset, we consider the lower bound on the number of sparse blocks. We first filter out all sparse blocks both datasets. We then geolocate the blocks and plot the percentage increase in the number of nonsparse blocks per country (Fig. 8a). The number of nonsparse blocks in Yemen, Suriname and Ethiopia increase from 17 each to, 713, 221, and 81 respectively. While the countries with the largest percent increase in number of nonsparse blocks lie outside of North America and Europe, US, France, UK, Australia and Brazil see the largest improvement in the raw number of nonsparse blocks.



(a) (b)

Fig. 8: (a) Improvement percentage in nonsparse block coverage over ANT by country. (b) Improvement percentage in reliable /24 block coverage over ANT by country.

Geolocating reliable blocks. We geolocate each /24 block in the combined dataset and then group the blocks by country. We then filter out the blocks with $A(E(b)) < 0.3$. Fig. 8b shows the per country percentage increase in the number of reliable /24 blocks in the combined dataset. The number of reliable blocks

mapped to French Polynesia, Suriname and Ethiopia increase from 37, 25 and 26, to 237, 230 and 208 respectively (Table 3). The baseline coverage in these countries however varies widely. For example, coverage in Ethiopia increased by 184 /24 blocks, and in Ireland by 19k /24 blocks.

Takeaway. Combining the two datasets produces relative increase in nonsparse and reliable /24 blocks in some countries outside North America and Europe.

8 Limitations and Future Work

This work is a step towards incorporating TCP and UDP-based probing into outage detection systems. We investigated the potential improvement in the coverage of /24 blocks for outage detection when using a combination of ICMP probes and other protocols. Specifically, we demonstrated that it is possible to reduce the number of sparse blocks and improve block availability for outage detection. Our method uses snapshots of data collected on different days and therefore, there might be some turnaround of hosts. We note that the initial availability scores in Trinocular are derived in a similar manner.

IP churn has been reported to be as high as 25% (observed from CDN logs) [20], and our analysis is based on point-in-time estimates of /24 block availability. We have merely shown that the initial estimation of block availability can improve with the addition of Censys data. There is still work to do in order to realize the actual gains in active monitoring of the IPv4 address space.

In this work, we consider probing a host using the protocol that yields the most responses for each IP. However, hosts may change IP addresses or change which services they are running. In such cases, probing via multiple protocols could potentially increase the likelihood of eliciting a response. Further analysis of our dataset could shed additional light on the most effective, yet practical, probing policies.

Based on the potential benefits described in this work, in future work, we plan to integrate TCP/UDP scanning techniques into the active probing signal of an operational outage detection system (IODA). Integrating these techniques into an operational outage detection system will also allow for more in-depth analysis of the degree to which additional probing techniques can improve the system's ability to outage detection. Extending IODA to support the additional probing techniques will require: getting consistent access to or conducting TCP/UDP scans of the full IPv4 address space across many ports (e.g., using ZMap), automating the analysis these scans to create a hit list of IPs and ports to monitor (i.e., applying the methodology of this work), integrating these hit lists into IODA's active measurement infrastructure, adding supporting for launching TCP/UDP probes, and integrating the results into IODA's dashboards.

Acknowledgements

We thank the anonymous reviewers for their thoughtful feedback and our shepherd, Dr. Kyle Schomp, for helping improve the presentation of this paper.

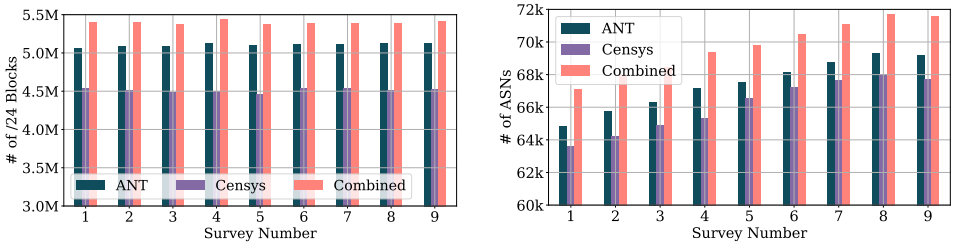
References

1. Ant datasets. https://ant.isi.edu/datasets/index.html (2022)
2. IODA. https://ioda.live (2022)
3. IPInfo. https://ipinfo.io/products/ip-database-download (2022)
4. RIPE Atlas. https://atlas.ripe.net/ (2022)
5. Baltra, G., Heidemann, J.: Improving coverage of internet outage detection in sparse blocks. In: Passive and Active Measurement: 21st International Conference, PAM 2020, Eugene, Oregon, USA, March 30–31, 2020, Proceedings 21. pp. 19–36. Springer (2020)
6. Bano, S., Richter, P., Javed, M., Sundaresan, S., Durumeric, Z., Murdoch, S.J., Mortier, R., Paxson, V.: Scanning the internet for liveness. ACM SIGCOMM Computer Communication Review **48**(2), 2–9 (2018)
7. CAIDA: pyipmeta. https://github.com/CAIDA/pyipmeta (2022)
8. Durumeric, Z., Adrian, D., Mirian, A., Bailey, M., Halderman, J.A.: A search engine backed by internet-wide scanning. In: Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security. pp. 542–553 (2015)
9. Durumeric, Z., Wustrow, E., Halderman, J.A.: ZMap: Fast Internet-wide scanning and its security applications. In: 22nd USENIX Security Symposium (USENIX Security 13). pp. 605–620 (2013)
10. Fan, X., Heidemann, J.: Selecting representative IP addresses for Internet topology studies. In: Proc. of IMC (2010)
11. Guillot, A., Fontugne, R., Winter, P., Merindol, P., King, A., Dainotti, A., Pelsser, C.: Chocolatine: Outage detection for internet background radiation. In: 2019 Network Traffic Measurement and Analysis Conference (TMA). pp. 1–8 (2019). https://doi.org/10.23919/TMA.2019.8784607
12. Heidemann, J., Pradkin, Y., Govindan, R., Papadopoulos, C., Bartlett, G., Bannister, J.: Census and survey of the visible internet. In: Proceedings of the 8th ACM SIGCOMM conference on Internet measurement. pp. 169–182 (2008)
13. Internet Addresses IPv4 Response History Dataset: Dataset, PREDICT ID: USC-LANDER/internet_address_history_it95w-20210727/rev12156. Provided by the USC/LANDER project. http://www.isi.edu/ant/lander (2022)
14. Katz-Bassett, E., Madhyastha, H.V., John, J.P., Wetherall, D., homas Anderson: Studying black holes in the internet with hubble. In: 5th USENIX Symposium on Networked Systems Design and Implementation (NSDI 08). USENIX Association, San Francisco, CA (Apr 2008), https://www.usenix.org/conference/nsdi-08/studying-black-holes-internet-hubble
15. Leonard, D., Loguinov, D.: Demystifying service discovery: implementing an internet-wide scanner. In: Proceedings of the 10th ACM SIGCOMM conference on Internet measurement. pp. 109–122 (2010)
16. NRO: RIR Statistics. https://www.nro.net/about/rirs/statistics/ (2023)
17. Padmanabhan, R., Schulman, A., Dainotti, A., Levin, D., Spring, N.: How to find correlated internet failures. In: Passive and Active Measurement: 20th International Conference, PAM 2019, Puerto Varas, Chile, March 27–29, 2019, Proceedings 20. pp. 210–227. Springer (2019)
18. Quan, L., Heidemann, J., Pradkin, Y.: Trinocular: Understanding Internet reliability through adaptive probing. SIGCOMM Comput. Commun. Rev. **43**(4), 255–266 (Aug 2013)
19. Richter, P., Padmanabhan, R., Spring, N., Berger, A., Clark, D.: Advancing the art of internet edge outage detection. In: Proceedings of the Internet Measurement Conference 2018. p. 350–363. IMC '18, Association for Computing Ma-

chinery, New York, NY, USA (2018). https://doi.org/10.1145/3278532.3278563, https://doi.org/10.1145/3278532.3278563
20. Richter, P., Smaragdakis, G., Plonka, D., Berger, A.: Beyond counting: new perspectives on the active IPv4 address space. In: Proc. of IMC (2016)
21. Schulman, A., Spring, N.: Pingin'in the rain. In: Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference. pp. 19–28 (2011)
22. Shah, A., Fontugne, R., Aben, E., Pelsser, C., Bush, R.: Disco: Fast, good, and cheap outage detection. In: 2017 Network Traffic Measurement and Analysis Conference (TMA). pp. 1–9. IEEE (2017)
23. Syamkumar, M., Gullapalli, Y., Tang, W., Barford, P., Sommers, J.: Bigben: Telemetry processing for internet-wide event monitoring. IEEE Transactions on Network and Service Management **19**(3), 2625–2638 (2022)

A Appendix 1

/24 blocks covered in the datasets. We count the number of active /24 blocks (/24 block with at least one responding host) in every snapshot, for both datasets and find the union of /24 blocks (Fig. 9a). On average, the coverage of /24 blocks increases by 293k when the ICMP probing is supplemented with scan data from Censys.



(a) **(b)**

Fig. 9: (a) /24 coverage across surveys from both datasets. On average, ANT covers 5.1 million /24 blocks and Censys-best covers 4.5 million /24 blocks. The combination of the two datasets yields about 5.39 million /24 blocks on average per survey (b) Autonomous systems identified in the two datasets. On average ANT scans span over 67k ASes and Censys over 66k ASes. The combination of the two datasets yields 69k ASes

ASes covered in the datasets. We look at the number of ASes found in each snapshot by both methods in Fig. 9b. On average, we find a 91% overlap in the ASNs between the two datasets. The set of ASes covered by Censys is slightly more than those covered by ICMP probing in the more recent surveys.

Survey dates for the datasets The scan/survey dates for ANT and Censys datasets are listed in Table 2.

Geolocation of reliable blocks. While countries like Yemen and Suriname see relative improvements in /24 coverage (Table 3), a large fraction of improvement in the number of reliable blocks are observed in North America and Europe.

Snapshot	ANT date	Censys date	Difference (days)
1	2020-11-10	2020-11-10	0
2	2021-03-06	2021-03-09	3
3	2021-06-19	2021-06-22	3
4	2021-08-27	2021-08-23	4
5	2022-01-28	2022-01-25	3
6	2022-04-19	2022-04-19	0
7	2022-06-29	2022-06-28	1
8	2022-09-09	2022-09-06	3
9	2022-12-03	2022-12-06	3

Table 2: Snapshot dates for ANT ISI and Censys datasets

Country	% increase in no. of reliable /24 blocks	Increase in number of blocks
Suriname	89.1	205
Ethiopia	87.5	182
French Polynesia	84.38	200
Andorra	80.55	145
Ireland	73.95	26,450

Table 3: Countries with the largest increase in % of reliable /24 blocks.

Table 4 lists the countries with the largest improvement to the number of reliable blocks. Table 4 shows the countries with the biggest percentage increase in number of reliable /24 blocks, for each continent.

Country	% increase in no. of reliable /24 blocks	Increase in number of blocks
United States	19.95	235071
Japan	22.05	42,679
Germany	21.38	39,718
France	34.97	391,12
Brazil	24.87	37,051

Table 4: Countries with the largest increase in number of reliable /24 blocks.