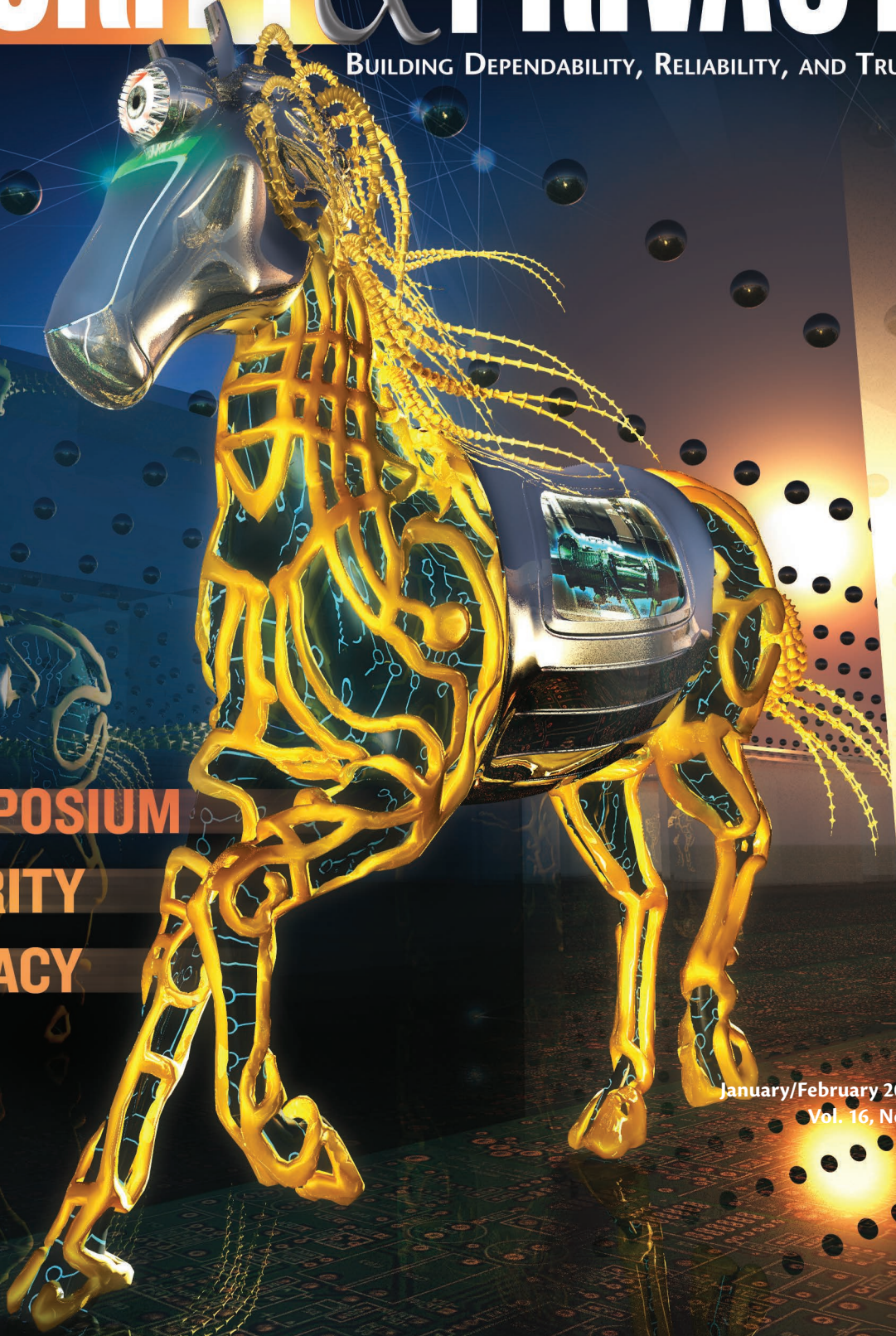


E-voting in Practice ■ Botnet Defense ■ Freedom of Encryption ■ Analyzing Flow Data

IEEE

SECURITY & PRIVACY

BUILDING DEPENDABILITY, RELIABILITY, AND TRUST



IEEE SYMPOSIUM
ON SECURITY
AND PRIVACY

January/February 2018
Vol. 16, No. 1





Toward Continual Measurement of Global Network-Level Censorship

Paul Pearce | University of California, Berkeley

Roya Ensafi | University of Michigan

Frank Li | University of California, Berkeley

Nick Feamster | Princeton University

Vern Paxson | University of California, Berkeley

Many accounts of censorship begin—and end—with anecdotes or short-term studies from a handful of vantage points. To enable continuous measurement of network-level censorship on an Internet-wide basis, Augur leverages TCP/IP side channels to measure reachability between two Internet locations from off-path third-party vantage points.

Ancedotes, news reports, and policy briefings collectively suggest that Internet censorship practices are pervasive.¹ Many countries employ a variety of techniques to bar their citizenry from accessing a wide spectrum of information and services, spanning the range from content sensitive for political or religious reasons, to microblogging, gambling, and pornography, to the use of censorship circumvention systems themselves.

Unfortunately our understanding of global censorship practices and their longitudinal trends is limited, despite censorship impacting literally billions of people. These limitations arise because we primarily derive our understanding of censorship practices and techniques from sparse case studies and accounts, which often heavily focus on the state of censorship in a single country as seen at a single point in time. We lack global views that comprehensively span the worldwide Internet, and we lack reliable continuous views that flag the onset of new censorship and relaxation of existing policies.

Existing approaches to collect continuous global censorship measurements include making use of

network proxies (for instance, ICLab) or in-country volunteers running mobile applications (for instance, OONI), and opportunistically leveraging user visits to instrumented websites.² These approaches remain difficult to deploy in practice: for example, some countries might not have volunteers or globally available VPN exits within them, or censors may already block the network access required for taking measurements. In addition, volunteer-based approaches can potentially implicate users who attempt to access prohibited Internet sites.

To address these limitations, we have developed Augur,³ a method for robustly inferring TCP/IP network disruptions between Internet endpoints and web services around the world from a single independent off-path location, using existing network protocol side channels. (For a more in depth treatment of our work, see “Augur: Internet-Wide Detection of Connectivity Disruptions.”³) Augur allows us to *continuously* gather measurements of network-level censorship from locations across the Internet and around the world without

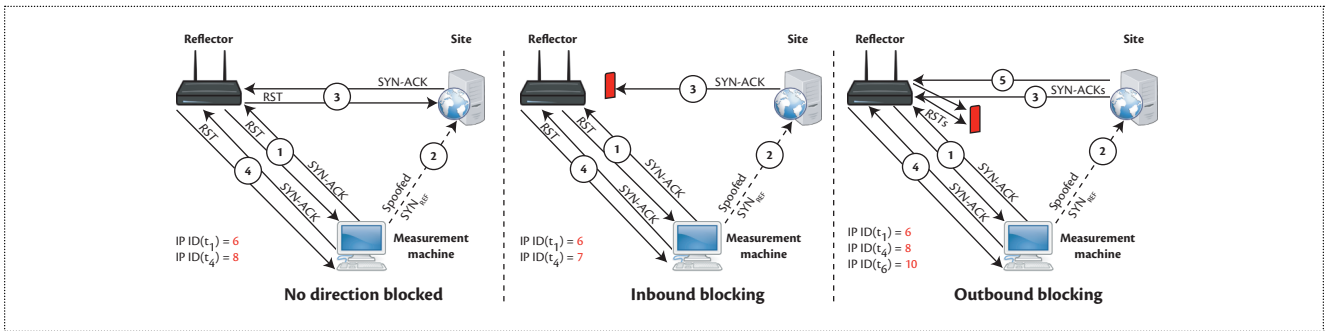


Figure 1. Overview of the basic method of probing and perturbing the IP ID side channel to identify potential censorship. “Reflectors” are Internet systems with a global IP ID counter, providing the side channel necessary for our inference. “Sites” are potentially blocked hosts that respond to TCP SYN packets on port 80. (In the right-hand figure, we omit subsequent measuring of the reflector’s IP ID by the measurement machine at time t_6 .) The spoofed SYN packets we transmit have a source field set to the reflector.

relying on volunteers or network proxies. To avoid potentially implicating individuals in such measurements, we also identify a process for finding infrastructure endpoints that are arguably safe to measure, as these systems are not in general linked to users.

We validate and demonstrate the potential of Augur through measurements of disruption Internet-wide across 179 countries and dependent territories over 17 days, using sensitive domains compiled by Internet censorship experts. We find that our results are consistent with prior smaller-scale or regional studies, as well as with expected filtering behavior. We also identify the top countries that experience connectivity disruption, highlighting many of the world’s most infamous Internet censors.

Inferring Network-Level Censorship

To gain global visibility into censorship behaviors, we need a method for gathering network measurements continuously from vantages in countries worldwide. Unfortunately, relying on network proxies, volunteers, or website visitors to collect measurements is neither reliable for continuous monitoring nor truly global in scale. As an alternative approach, we employed a previously developed method⁴ for inferring TCP/IP connectivity between two hosts (such as a router and a website’s server) from an independent third vantage point. This technique relies on protocol *side-channel information*: externally visible information that is dependent on, and hence can reveal, the internal state of network communication.

Side channels occur when an action has an unintended effect that manifests to a careful observer. For example, say a fast food restaurant decides to assign order numbers sequentially. If someone orders from the restaurant twice within a day, they can estimate the amount of business at the restaurant by comparing their two order numbers. Although the order number is for

uniquely associating a food order with a customer, it serves as a side channel by revealing more information than intended.

Augur relies on a network side channel in the headers of IP packets, specifically the IP identifier (IP ID) field. The IP ID is a 16-bit field intended to aid in the reconstruction of fragmented packets. In practice, many Internet hosts assign IP ID values using a single global counter that increments by 1 with each IP packet sent, regardless of whether the packets belong to the same flow. The change in such a host’s counter over time reflects the number of packets that host has sent, and is visible to any party it communicates with. Thus, we can use this IP ID side-channel information to infer whether a host is generating IP packets.

Figure 1 presents an overview of how we can use IP ID side channels to identify several kinds of network disruptions. To probe the IP ID value of some host over time, a third-party vantage point we call the *measurement machine* sends specially crafted response packets (TCP SYN-ACK) to the host. As the host never sent an initial packet, these response packets will cause the host to generate error packets (TCP RSTs) in reply. The measurement machine monitors these error packets to track the evolution of the host’s IP ID global counter. We monitor the IP ID values of a host *reflector*, a term denoting that the host reflects RST packets from both our measurement machine and (subsequently) from an endpoint that a censor may be trying to filter. The reflector is a host in a network that may experience network-level blocking.

We call the other endpoint of this connection the *site*. The site represents a potential target of censorship, that is, a remote resource or service that a censor operating in the reflector’s country may seek to block. Here we use web servers operating on port 80 (HTTP) as the site, although Augur can measure blocking of any TCP-based service.

To test for network connectivity, we must induce traffic between the reflector and the site. To do so, we will need to inject incorrect (“spoofed”) traffic into the network. In this process, the measurement machine sends a connection request (TCP SYN) packet to the site with its IP source address falsely set to that of the reflector. Thus, the site believes that the reflector wishes to initiate a connection, and responds with a SYN-ACK packet. In the event of network censorship, several different scenarios may follow:

- *No direction blocked.* If the site’s response (SYN-ACK) reaches the reflector, the reflector emits an error (RST) packet, as it did not initiate a connection. This results in a single increment of the reflector’s IP ID counter between measurements by 1, a change visible via our probing. If the RST reaches the site, then the site will delete its nascent TCP connection with the reflector, and take no further action. This case reflects a situation without any censorship, as TCP/IP packets can be successfully delivered in both directions.
- *Inbound blocking.* A censor may block the site’s SYN-ACK from reaching the reflector. In this case, the reflector does not generate any traffic in response, and thus its IP ID global counter does not increment between our probes.
- *Outbound blocking.* A censor may allow inbound response packets, but block the reflector’s outbound RSTs sent to the site. In this case, because the site never receives RST, its TCP implementation will assume that the earlier SYN-ACK may have been dropped while in transit. Most implementations will consequently retransmit the SYN-ACK multiple times (often at least three times), each of which elicits an RST from the reflector, but the RST never makes it to the site due to the censor’s blocking. However, each of the elicited RSTs will increment the reflector’s global IP ID counter, resulting in a total increase greater than 1 between our probes.

Thus, by tracking the evolution of a reflector’s IP ID global counter as we induce traffic between it and a site, we can infer censorship activity based on to what degree, and when, the IP ID counter changes.

Unfortunately, this side channel is noisy; the previous example assumes ideal conditions where the reflector communicates only with the site. However, hosts communicate with other machines on the Internet, and packets can be lost. These factors make precise inferences of IP ID counters difficult, as counters will change independent of our measurements. To address this noise, Augur uses repeated measurements and the statistical technique of *sequential hypothesis testing* (SHT) to probabilistically identify disruption with a specified

degree of confidence. We describe this method further below. In addition, this technique requires that both reflectors and sites exhibit specific properties described further in “Augur: Internet-Wide Detection of Connectivity Disruptions.”³

Inferring Network-Level Censorship in the Presence of Noise

The distillation of our method to detect connectivity disruptions involves introducing potential perturbations in an IP ID counter at the reflector, and subsequently observing whether this value’s evolution reflects the attempted perturbations. Unfortunately, the noise associated with both the counter and its measurement necessitate finding ways to extract reliable signals.

Approach: Statistical Detection

We periodically measure the natural evolution of a reflector’s counter in the absence of perturbation as a control that we can compare against the evolution of the IP ID under perturbation. We then attempt to perturb the IP ID counter by inducing network traffic between the reflector and the site, and subsequently measure the evolution of this counter. We take care not to involve any site or reflector in multiple simultaneous measurements, since doing so could conflate two distinct results.

Ultimately, we are interested in detecting whether the IP ID evolution for a reflector changes as a result of the perturbations we introduce. We can represent this question as a classical problem in statistical detection. In designing this detection method, we first must determine the measure that captures IP ID evolution, and we represent its distribution of values with a random variable Y . We also must decide on the specific detection approach that allows us to distinguish the distribution of values for Y when the event does and does not occur. We choose IP ID *acceleration* (that is, the second derivative of IP ID between successive measurements) as our measure as it ideally has a zero mean, regardless of reflector. With a zero mean, the distribution of the random variable should be stationary and should be similar across reflectors. Conceptually, we can think of reflectors, at arbitrary times, being equally likely to experience an increase in traffic rate as a decrease.

Detection Framework: Sequential Hypothesis Testing

We use sequential hypothesis testing for the detection algorithm. SHT is a statistical framework for real-time decision making, where the decision concerns which of two possible models (that is, underlying situations) best explains a sequence of observations. The framework operates in terms of an observable random Bernoulli variable for which the two models assign differing

probabilities (“priors”) for the true/false outcomes. The framework takes the probabilities for each prior along with tolerable false positive and negative rates as input, and for repeated trials continually updates the probability of each model explaining the observations. This process continues until the observations allow making a decision preferring one model over the other prior with the specified false positive and negative rates. SHT’s ability to perform online detection subject to tunable false positive/negative rates, and its tolerance to noise, makes it well-suited to our detection task. In addition, it is possible to compute an expectation for the number of trials required to produce a detection, thus enabling efficient measurement.

Figure 2 illustrates the SHT detection algorithm, which performs a series of sequential hypothesis tests to detect possible inbound blocking. (A similar construction extends this approach to detecting outbound blocking.) As we observe each trial, we update the likelihood ratio function $\Lambda(Y)$ based on the prior probabilities. Once updated, we compare the value of $\Lambda(Y)$ against the thresholds η_0 and η_1 , bounds derived from the prior probabilities. If $\Lambda(Y) \leq \eta_0$, we accept the evidence as reflecting the presence of inbound blocking.

If $\Lambda(Y) \geq \eta_1$, we conclude that IP ID acceleration occurred as a result of no inbound blocking. This does not give us a final result, as we still must decide between outbound blocking and no blocking. To make this decision, we proceed to another SHT phase, which we omit here for brevity.

A third possible output of the algorithm is that $\Lambda(Y)$ did not meet either threshold. If we can conduct more trials, we restart the algorithm. If not, we output that blockage is unknown.

Expected Number of Trials

The SHT framework also calculates the expected number of trials needed to arrive at a decision for inbound and outbound blocking, as a function of the prior probabilities and acceptable false positive and negative rates. A majority of reflectors require fewer than 10 trials to identify inbound blocking, and fewer than 20 trials to identify outbound blocking.

False Positives and Negatives

P_F is defined as the false positive probability, and P_D as the detection probability. The complement of the detection probability, $1 - P_D$, is the probability of false negatives. These values express the probability of a false result for a single SHT experiment (set of trials). However, for our method, we perform numerous SHT experiments across sites and reflectors. To account for these repeated trials, we set both P_F and $1 - P_D = 10^{-5}$. This value results in the expectation of less than one

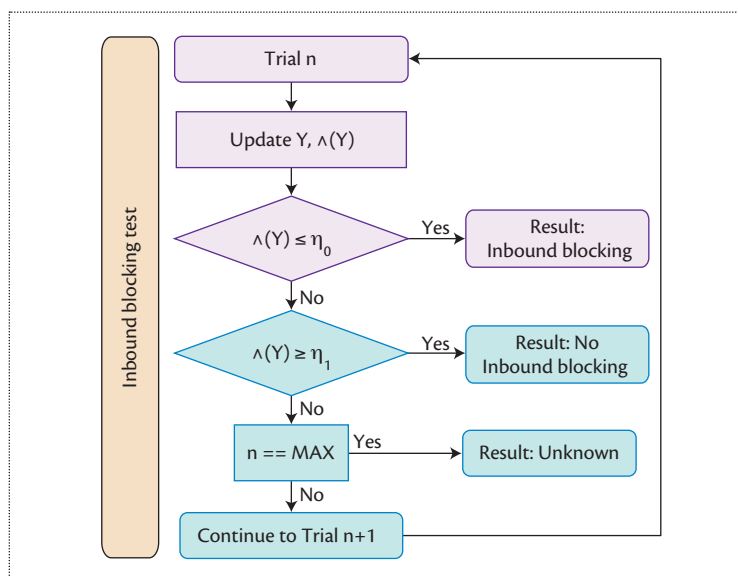


Figure 2. Flow chart of our algorithm to identify inbound blocking using a series of sequential hypothesis tests. We use a similar construction to test for possible outbound blocking. SHT provides the means to compute η_0 and η_1 from priors associated with the different potential outcomes.

incorrectly classified site per reflector. Given that as P_F and $1 - P_D$ decrease, the expected number of trials to reach a decision increases, and our selection of a small value negatively affects our ability to make decisions. This effect is somewhat mitigated by the distance we find in practice between empirically derived priors.

Ethics

The side-channel measurement method we develop induces traffic between the reflector and the site, a potentially censored destination. An inexperienced or imprecise observer of these network measurements may (wrongly) conclude that the person who operates or owns the reflector was willfully accessing the site. However, they can properly distinguish Augur’s activities by observing that no TCP connection is actually established (the TCP three-way handshake is not completed). Fundamentally, Augur increases the risk that someone residing in a censored region will be blamed for contacting a disallowed site. The additional risks introduced by such measurements are not fully known, and they vary by country; the risks are also continually evolving, which further complicates risk assessment.

We guide our design decisions according to the ethical reasoning framework of the Menlo⁵ and Belmont Reports,⁶ which outline several principles:

- *respect for humans* (limiting the potential harm to humans),

- *beneficence* (weighing the benefits against the risks of the experiment),
- *justice* (considering whether those bearing the risks of the experiment are also those who may benefit in some way), and
- *respect for law and public interest*.

Ethicists have explored these considerations, as well as ethical approaches one can take to minimize additional risk when obtaining informed consent is not feasible or practical (as in our case). Importantly, these principles may in general sometimes be in conflict, so they should be viewed more as a reasoning framework than as a checklist.

We design our measurement methods to minimize the potential harm and additional risk to humans by selecting each reflector in a way that minimizes the likelihood that the resulting measurement traffic could be associated with a human. To do so, we select reflectors that very likely correspond to Internet infrastructure (for example, internal routers and middleboxes), as opposed to hosts that belong to individual citizens (for example, laptops, desktops, and home routers).

To identify suitable Internet infrastructure, we use the CAIDA Ark dataset. ARK contains network path (traceroute) measurements to one randomly selected IP address in every possible /24 IPv4 prefix (most significant 24 bits). We include a reflector in our experiments only if it appears in an Ark traceroute at least two hops away from the traceroute endpoint, as this likely is out of the end user's local network.

Although this approach increases the likelihood that the reflector IP addresses are routers or middleboxes, the method is not foolproof. Devices that are attributable to individuals might still be two hops from the network edge, or a network operator might be held accountable for the perceived actions performed by their machines. Our techniques do not eliminate the increase in risk. Rather, they reduce it to the point where the benefits of collecting these measurements arguably outweigh the additional risks of collecting them.

Our measurements send packets to various hosts ("scanning"). We follow the guidelines for ethical scanning behavior outlined by Durumeric and colleagues,⁷ which also abide by the principle of respect for law and public interest. We limit our scanning rate to reduce undue traffic load to Internet-connected devices, and we couple our measurements with visible indicators that clearly tag our measurements as being related to research purposes. We also respect any requests to opt out of our scanning.

Deploying Augur

To understand the potential for our approach to continuously identify Internet censorship, we deployed Augur

and conducted an Internet-wide measurement study. In this section, we discuss the details of our deployment as well as how we validated our results.

Reflector Selection

The first step in finding reflectors requires us to find hosts on the Internet that generate RSTs in response to unsolicited SYN-ACKs. To find these hosts, we developed a new probe module for ZMap,⁷ an Internet-scale network scanner capable of efficiently scanning the entire IPv4 address space. Our probe module first sends SYN-ACK packets to port 80 (HTTP) to hosts across the Internet and looks for well-formed RST responses. For candidate reflectors that do generate valid RSTs, we then must determine if they also have a global IP ID counter. To check this property, we perform several rounds of further probing to monitor the IP ID in each RST. The set of experimentally viable reflectors are then limited to those that represent Internet infrastructure using Ark data, due to ethical considerations. For our study, we randomly sampled a subset of reflectors within each country, geolocating each reflector's country using the MaxMind geolocation service.

Site Selection

We investigated websites historically observed as censored, as well as popular websites. For sensitive censored sites, we used CLBL, a list of 1,210 websites curated by the civil society group Citizen Lab.⁸ For popular sites, we randomly sampled 1,000 domains from the Alexa top 10,000 sites. Due to some overlap between the popular and frequently censored websites, in total we selected 2,134 websites. For each site, we resolved its domain name to identify the corresponding IP addresses.

Measurement Dataset

Table 1 summarizes our dataset and its Internet-wide geographic diversity. Across the entire Internet, we identified 23M viable reflectors in 234 countries and dependent territories. Applying our ethical framing and limiting hosts to Internet infrastructure reduced the set of viable reflectors to 53K across 179 countries and dependent territories. We then selected a random subset of reflectors per country, yielding a final experiment set of 2,050 across 179 countries. Using these 2,134 sites and 2,050 reflectors, we conducted 182M connectivity disruption network trials over 17 days, using the experiment we describe next. Our symposium paper discusses further details, such as how we filter out problematic sites and reflectors.³

Experiment Setup

Each SHT trial was composed of a collection of one-second time interval measurements of the network

Table 1. Summary of our reflector datasets, including the geographic diversity by the number of countries/dependent territories. All viable reflectors came from the IPv4 address space; ethically usable ones came in addition from a random subset of routers at least two hops away from traceroute endpoints in the Ark data.

Reflector datasets	Total reflectors	Number of countries
All viable	22,680,577	234
Ethically usable	53,130	179
Experiment sample	2,050	179

connectivity between an individual site and reflector. For each time interval, we measured the IP ID state of the reflector independent of all other tasks. At the start of each trial, we performed a number of calibration steps that ensured the site was up and functioning normally. We then waited four seconds before injecting spoofed SYNs toward the site, as shown in Figure 1. We used the reflector measurements before and after injection to compute the respective control and injection prior probabilities in our SHT formulation for inbound blocking detection. (We used a similar construction for identifying outbound blocking.) At the end of each trial, we performed a number of correctness and safety checks to ensure that both the site and reflector were online and operating correctly.

Validation

The utility of our method ultimately rests on its ability to accurately assess potential connectivity disruptions using a large number of measurement vantage points. Validating its findings presents significant challenges, as we lack comprehensive ground truth. In these circumstances, the best we can do is to analyze the aggregate results produced and confirm that they correspond to reasonable expectations about the employment of connectivity disruption.

More concretely, one would expect the set of sites disrupted by a network censor to be biased toward sites that are known to be sensitive and experience censorship. From this notion, we can examine the set of sites blocked by each reflector and ask how that population compares to the measurement population. This measure does not guarantee correctness, but it increases confidence in the observations given the inability to obtain ground truth.

Figure 3 shows, in aggregate, the bias of connectivity disruption toward commonly censored websites. About 57 percent of websites in the input site dataset came from the CLBL, demarcated in the plot with a vertical dotted line (the “CLBL bias line”).

If a blockage we observed reflects phenomena independent of censorship, we would expect to find that roughly 57 percent of sites blocked by reflectors to come from the CLBL. The results, however, show a considerable skew toward CLBL sites for both inbound and outbound blocking. We see this with the bulk of histogram volume lying to the right of the vertical dotted CLBL bias line. Excluding reflectors with fewer than five blocked sites to avoid small number effects, we observe that for 99 percent of reflectors, more than 57 percent of inbound filtering involves CLBL sites. Similarly, we find 95 percent of outbound filtering biased toward the CLBL. These observed biases agree with our prior expectations that we should find CLBL sites more widely censored.

Country-Level Network Censorship

The accordance of our results with prior small-scale case studies on censorship in particular countries provides further confidence in Augur’s accuracy, and demonstrates the potential of such side-channel methods.

To examine blocking by country, we use MaxMind’s geolocation services to identify each reflector’s country, then aggregate all reflectors within each country to compute the mean percentage of site blocking. This approach ultimately has two fundamental limitations. First, errors in MaxMind’s reflector geolocation may place a reflector in the wrong country, influencing and possibly biasing our results. Second, comparative results depend on which domains are measured. If our dataset of sensitive sites is biased toward certain countries (that is, has more content from those countries), those countries may appear more censored. Performing comparative censorship studies resistant to geolocation errors and biased domain datasets remains an open research problem in need of future investigation.

Table 2 lists the 10 countries with the largest percentages of blocked sites as seen by at least one reflector in a given country, along with the country’s bias toward content on the CLBL. Figure 4 portrays this at a global

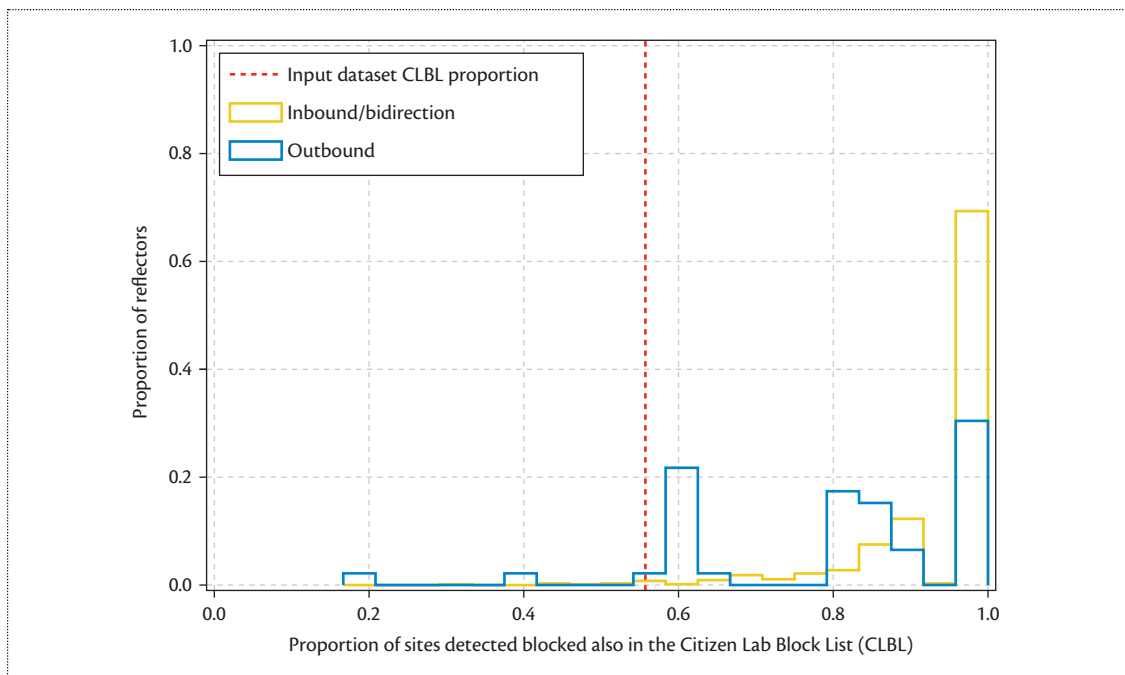


Figure 3. Histogram of bias of blocked sites toward CLBL sites. CLBL sites consist of 56.7 percent of our sites, demarcated at the dotted vertical line. Reflectors to the right of the demarcation line have a blocking bias toward known sensitive sites.

Table 2. Summary of the 10 countries with the highest percentage of blocked sites in our dataset.

Country	Block (%)	CLBL (%)	Number
China	5.0	70.9	36
Iran	3.4	55.7	14
Sudan	2.2	54.3	12
Russia	1.8	78.9	17
Latvia	1.8	81.6	14
Turkey	1.8	83.8	15
Hong Kong	1.7	88.9	16
Colombia	1.7	85.7	16
Libya	1.5	77.4	10
United Kingdom	1.4	90.0	16

scale, illustrating that some degree of connectivity disruption is experienced by hosts geolocated in countries around the world.

We see that many of the countries manifesting the most disruption correspond closely with countries

known to heavily censor, such as China, Iran, Sudan, Russia, and Turkey.⁹ (We list Hong Kong separately from China, although traffic from Hong Kong may traverse Chinese networks and experience disruption.) Of the 10 countries with the highest average blocking, the Open-Net Initiative⁹ has reported Internet censorship of political or social material in every country except Latvia and the United Kingdom. More recently, reports have documented Latvia as heavily censoring gambling websites and political content.¹⁰ Our results appear plausible for the United Kingdom as well, which has a history of filtering streaming and torrent sites¹¹ and adult content.¹²

These disruptions may actually be implemented in different ways within a single country. If so, the differences result in non-uniform filtering policies, as has been observed with the Great Firewall of China¹³ and UK adult content filtering.¹² We observe that for most countries, there exists some variation in the disruption experienced by reflectors within a country, suggesting that interference indeed often differs across networks even within a country. We find this behavior extends widely, highlighting the importance of connectivity measurements from numerous vantage points within a country, since findings may differ across nearby networks and locations.

Discussion

Here we discuss some of the limitations of our current measurement approach, and the types of questions

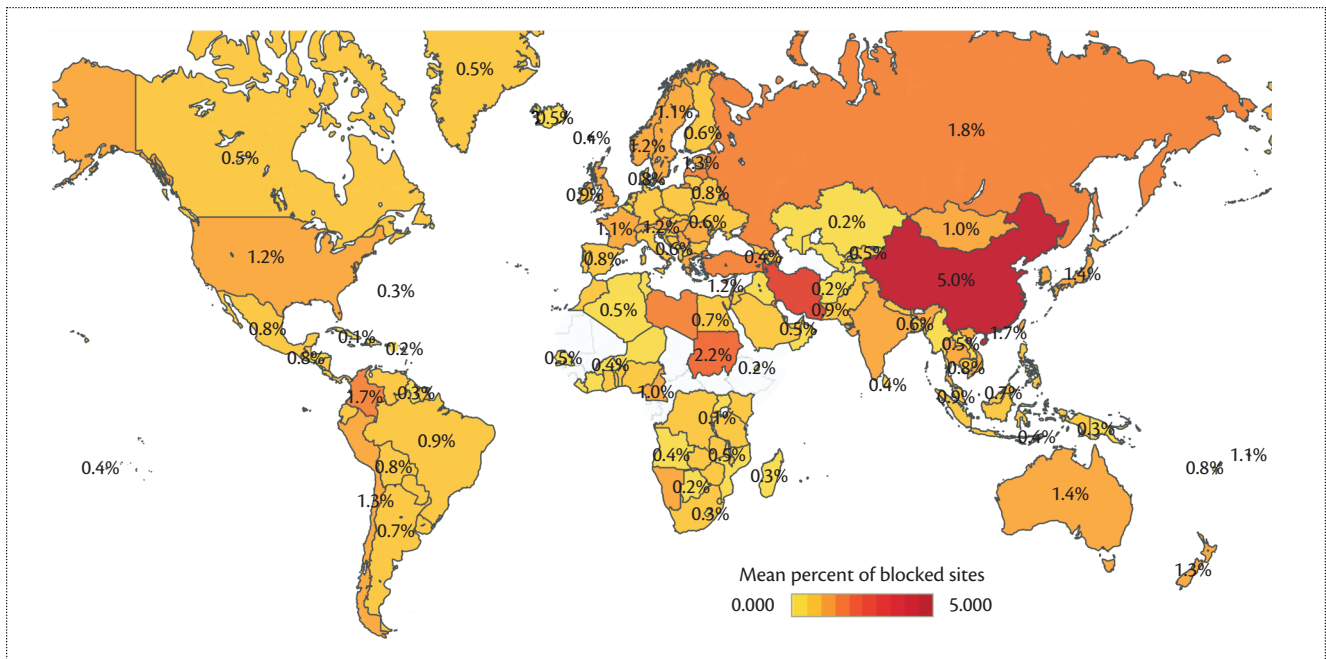


Figure 4. Global heat map showing the percentage of measured sites filtered for any reflector in countries around the world. Relative filtering amounts are influenced by the composition of sites measured. China experiences the highest average amount of our dataset filtered, at 5 percent of sites filtered by resolvers within the country. In countries such as the United States, this filtering can be a product of voluntary censorship, such as the deployment of corporate filtering software. These phenomena require further study.

that a longitudinal system might help us answer going forward.

Limitations

Our current measurement methods have limitations concerning coverage, granularity, and location accuracy. First, our restriction on the reflectors we use for ethical purposes impacts our coverage, and more exhaustive techniques would increase the number of hosts we could measure with. Second, our measurements do not reveal where along an end-to-end path censorship might be occurring, preventing us from better understanding the expanse of filtering across hosts in an entire region. Third, existing IP geolocation tools have known inaccuracies,¹⁴ particularly for Internet infrastructure. Further work is necessary to understand the biases and limitations imposed by geolocation errors in comparative studies. As geolocation techniques improve, particularly for IP addresses that correspond to Internet infrastructure, we can develop more confidence in our country-level characterizations. Finally, network phenomena such as rerouting, traffic shaping, and transient network failures can make it difficult to disambiguate overt filtering actions from more benign network management practices. Future work should explore the effects of these phenomena on Augur’s measurement accuracy, as well as mitigations.

What Questions Can We Answer?

With a longitudinal, continuous view of Internet censorship using multiple techniques, we can begin to answer numerous questions previously difficult to address. These include:

- *How do Internet censorship trends vary over time?* Understanding how blocking changes over time allows us to identify events and actions that influence a censor’s behavior. These trends are critical, both qualitatively and quantitatively, for a range of social science research aimed at understanding and addressing censorship.
- *How do Internet censors respond to circumvention, and on what time scale?* Related to censorship trends is how censors respond to active circumvention of their efforts. Understanding how censors engage and on what time scales provides empirical grounding for evaluating various defenses and enables constructing more effective circumventions.
- *Does censorship vary between regions within countries?* Examining how censorship varies between regions provides insights into how censorship systems are deployed, both technically and organizationally. These insights can better inform circumvention and policy interventions. Augur’s use of multiple vantage points across a range of networks within a country

allows us to actively explore and understand these deployments.

The ability to provide comprehensive empirical footing for these questions can enable new lines of research as well as serve as an invaluable resource for social scientists.

The continuous, widespread measurements that we can collect with these techniques can complement anecdotes, news reports, and policy briefings to ensure that we can support future assessments of Internet filtering with sound, comprehensive data. Part of this transition to practice involves further developing the system that we have designed to facilitate ongoing operation, including automating the validation of the measurements that we collect and the correlation with other datasets and tools. Although the data that we collect reflects only TCP/IP-based filtering, Augur complements other methods we have developed to facilitate other types of censorship measurement, such as DNS manipulation.¹⁵ We aim to ultimately compare results produced by multiple methods, including datasets from volunteer-based measurement platforms. ■

Acknowledgments

We thank Randy Bush, Jed Crandall, David Fifield, Sarthak Grover, and Brad Karp. This work was supported in part by National Science Foundation Awards CNS-1237265, CNS-1518878, CNS-1518918, CNS-1540066, and CNS-1602399.

References

1. "Freedom on the Net," Freedom House, 2016; <https://freedomhouse.org/report/freedom-net/freedom-net-2016>.
2. S. Burnett and N. Feamster, "Encore: Lightweight Measurement of Web Censorship with Cross-Origin Requests," ACM SIGCOMM, 2015.
3. P. Pearce et al., "Augur: Internet-Wide Detection of Connectivity Disruptions," IEEE Symposium on Security and Privacy, 2017.
4. R. Ensafi et al., "Detecting Intentional Packet Drops on the Internet via TCP/IP Side Channels," Passive and Active Measurements Conference (PAM), Springer, 2014; <http://arxiv.org/abs/1312.5739>
5. D. Dittrich and E. Kenneally, *The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research*, tech. report, US Department of Homeland Security, Aug 2012.
6. "The Belmont Report—Ethical Principles and Guidelines for the Protection of Human Subjects of Research," US Department of Health and Human Services, 18 Apr. 1979; <https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/index.html>.
7. Z. Durumeric, E. Wustrow, and J.A. Halderman, "ZMap: Fast Internet-Wide Scanning and Its Security Applications," USENIX Security Symposium, 2013.
8. "Block Test List," Citizen Lab, <https://github.com/citizenlab/test-lists>.
9. "OpenNet Initiative," <https://opennet.net>.
10. A. Spence, "Russia Accuses Latvia of 'Blatant Censorship' after Sputnik News Site Is Shut Down," Politico, 30 Mar. 2016, www.politico.eu/blogs/on-media/2016/03/russia-accuses-latvia-of-blatant-censorship-after-sputnik-news-site-is-shut-down.
11. D. Bolton, "Putlocker Blocked in the UK by Internet Service Providers after High Court Order," Independent, 26 May 2016; <https://goo.gl/s8Hb43>.
12. S. Mitchell and B. Collins, "Porn Blocking: What the Big Four ISPs Actually Did," Alphr, 4 Aug. 2015; <http://www.alphr.com/networking/20643/porn-blocking-what-the-big-four-isps-actually-did>.
13. R. Ensafi et al., "Analyzing the Great Firewall of China over Space and Time," Privacy Enhancing Technologies (PETS), 2015.
14. B. Huffaker, M. Fomenkov, and k. claffy, *Geocompare: A Comparison of Public and Commercial Geolocation Databases*, tech. report, Cooperative Association for Internet Data Analysis (CAIDA), May 2011.
15. P. Pearce et al., "Global Measurement of DNS Manipulation," 26th USENIX Security Symposium (USENIX Security 17), USENIX Association, 2017.

Paul Pearce is a senior PhD student at the University of California, Berkeley, advised by Vern Paxson, and a member of the Center for Evidence-Based Security Research (CESR). His research brings empirical grounding to Internet security problems including censorship, cybercrime, and advanced persistent threats (APTs). Contact him at pearce@cs.berkeley.edu.

Roya Ensafi is a research assistant professor in computer science and engineering at the University of Michigan, where her research focuses on computer networking and security. She pioneered the use of side channels to remotely measure network interference and censorship of Internet traffic. Prior to joining the University of Michigan, she was a postdoc at Princeton University. Contact her at ensafi@umich.edu.

Frank Li is a PhD student at the University of California, Berkeley. His research mainly focuses on improving the remediation process for security issues such as vulnerabilities and misconfigurations. More broadly,

he is interested in large-scale network measurements and empirical studies in a computer security context. Contact him at frankli@cs.berkeley.edu.

Nick Feamster is a professor in the Computer Science Department at Princeton University and the deputy director of the Princeton University Center for Information Technology Policy (CITP). Before joining the faculty at Princeton, he was a professor in the School of Computer Science at Georgia Tech. He received his PhD in computer science from MIT in 2005. His research focuses on many aspects of computer networking and networked systems, with a focus on network operations, network security, and censorship-resistant communication systems. He is an ACM Fellow. Contact him at feamster@cs.princeton.edu.

Vern Paxson is a professor of electrical engineering and computer sciences at the University of California, Berkeley, and leads the Networking and Security Group at the International Computer Science Institute in Berkeley. His research focuses heavily on measurement-based analysis of network activity and Internet attacks. He works extensively on high-performance network monitoring, detection algorithms, cybercrime, and countering censorship and abusive surveillance. Contact him at vern@berkeley.edu.

myCS

Read your subscriptions through the myCS publications portal at <http://mycs.computer.org>




Take the CS Library wherever you go!

 IEEE Computer Society magazines and Transactions are now available to subscribers in the portable ePub format.

Just download the articles from the IEEE Computer Society Digital Library, and you can read them on any device that supports ePub. For more information, including a list of compatible devices, visit

www.computer.org/epub

 **IEEE** IEEE  computer society