Measuring the Deployment of Network Censorship Filters at Global Scale

Ram Sundara Raman[*], Adrian Stoll[*], Jakub Dalek[†], Armin Sarabi[*], Reethika Ramesh[*], Will Scott[‡], Roya Ensafi[*]
[*]University of Michigan, {ramaks, adrs, arsarabi, reethika, ensafi}@umich.edu
[†]The Citizen Lab, University of Toronto, jakub.dalek@utoronto.ca
[‡]Independent, willscott@gmail.com

***Abstract*—Content filtering technologies are often used for Internet censorship, but even as these technologies have become cheaper and easier to deploy, the censorship measurement community lacks a systematic approach to monitor their proliferation. Past research has focused on a handful of specific filtering technologies, each of which required cumbersome manual detective work to identify. Researchers and policymakers require a more comprehensive picture of the state and evolution of censorship based on content filtering in order to establish effective policies that protect Internet freedom.**

In this work, we present FilterMap, a novel framework that can scalably monitor content filtering technologies based on their blockpages. FilterMap first compiles in-network and new remote censorship measurement techniques to gather blockpages from filter deployments. We then show how the observed blockpages can be clustered, generating signatures for longitudinal tracking. FilterMap outputs a map of regions of address space in which the same blockpages appear (corresponding to filter deployments), and each unique blockpage is manually verified to avoid false positives.

By collecting and analyzing more than 379 million measurements from 45,000 vantage points against more than 18,000 sensitive test domains, we are able to identify filter deployments associated with 90 vendors and actors and observe filtering in 103 countries. We detect the use of commercial filtering technologies for censorship in 36 out of 48 countries labeled as 'Not Free' or 'Partly Free' by the Freedom House "Freedom on the Net" report [26]. The unrestricted transfer of content filtering technologies have led to high availability, low cost, and highly effective filtering techniques becoming easier to deploy and harder to circumvent. Identifying these filtering deployments highlights policy and corporate social responsibility issues, and adds accountability to filter manufacturers. Our continued publication of FilterMap data will help the international community track the scope, scale and evolution of content-based censorship.

I. Introduction

Governments and authorities increasingly seek to control how their citizens access content and communicate online, often citing concerns of national sovereignty, security, public morality, and terrorism. These information controls typically take the form of blocking access to certain websites or online services. While the technical means of censorship vary and continue to evolve, some of the most common forms include DNS tampering, injecting or dropping packets at the IP layer, and application-layer filtering using deep packet inspection (which enables advanced classification and filtering). For example, ISPs block sensitive keywords or prevent the use of "undesirable" applications (e.g., virtual private networks or Tor), sometimes by injecting a *blockpage*, a notice that explains to the user why the content has been made unavailable.

A worrying trend in recent years has been the proliferation of content filtering technologies, specialty networking products that inspect, filter, and/or tamper network communication for purposes other than packet forwarding. A decade ago, this filtering was expensive to conduct at the scale of a national network and its deployment was limited to a small number of motivated countries. Today, commoditization has brought carrier-grade content filtering within the budget of most governments, including those with poor human rights records [40].

Monitoring the use of filtering technologies for censorship can drive change in the regulation and behavior of companies selling filtering products. For instance, Citizen Lab [49] conducted investigations of a Canadian content filtering vendor Netsweeper showing how their products were employed in censorship systems around the world. The investigation was the result of several years of manual effort in identifying and scanning for product signatures. One particularly egregious case they identified concerned an "Alternative Lifestyles" blacklist curated by Netsweeper, which was used by several countries to block LGBTQ+ content. After advocacy based on Citizen Lab's findings, Netsweeper claims they have removed the option to block based on this category [57]. In another case, when ONI [41] informed Websense about the use of their product for censorship by the Yemeni government, they stopped providing software updates to the products deployed in Yemen [63].

Measuring the deployment of filtering technologies, which we refer to as *filters*, has been a cumbersome process as it involves manually identifying a unique *signature* for a small set of filter products, often while having physical access to a sample product and the assistance of on-the-ground collaborators, and then performing network scans using the signatures to detect deployments. Moreover, monitoring their deployment continuously requires sustainable systems. As a result, the censorship measurement community has only identified a handful of filters over the years. In 2013, Dalek et al. manually created signatures for *four filter vendors* and

Network and Distributed Systems Security (NDSS) Symposium 2020
23-26 February 2020, San Diego, CA, USA
ISBN 1-891562-61-4
https://dx.doi.org/10.14722/ndss.2020.23099
www.ndss-symposium.org

measured their deployment in several countries [15]. These signatures were based on the blockpages exposed by filters with public-facing IP addresses and were limited to a particular product configuration. In 2014, Jones et al. introduced a technique to reduce manual effort in finding filter signatures by clustering blockpages in ONI data [42] based on page length and term frequency vectors [34]. Unfortunately, the metrics were shown to be a poor heuristic for detecting blockpages, as the technique suffers from high false positives due to natural variations in page length from dynamic and language-specific content [62]. The inability to monitor the proliferation of filters more broadly withholds researchers, regulators, and citizens from efficiently discovering and responding to the misuse of these "dual-use technologies". This concern is echoed by a recent report to the UN Human Rights Council that addresses challenges in establishing regulatory measures and safeguards pertaining to the use of these products for censorship and surveillance [54].

In this paper, we present FilterMap, a framework for semi-automatically identifying filters that are configured to censor with user-observable blockpages, and we apply the technique to measure censorship filter deployments around the globe. FilterMap consists of two main phases: (*1*) **data collection**, in which blockpages are retrieved using network interference measurement techniques; and (*2*) **data analysis**, in which the collected data is processed to generate clusters of blockpages, each labeled by a unique signature, which identify filter deployments in different countries. We use iterative classification and image clustering to substantially automate the classification of injected responses. This automation represents a vital improvement, given the scale of the data collected (hundreds of millions of measurements).

A crucial challenge for our effort is to ethically collect a widespread and diverse set of blockpages that result from triggering many globally deployed censorship filters. Most censorship measurement techniques rely on volunteers or in-network vantage points that hinder performing measurements of the scale and frequency required to gather continuous data. Moreover, triggering these filters to act on forbidden or sensitive content can compromise user safety and requires a great deal of manual effort in surmounting language and cultural barriers to convey risks to volunteers [51]. Fortunately, recent measurement techniques [56], [46], [44], [43], [21] remove the need for an in-network volunteer and enable measurements to be performed remotely and safely across a broad swath of networks. We apply and extend one such technique, Quack [56], to investigate the global proliferation of censorship filters.

Using FilterMap we provide the first global-scale measurement of censorship filters and identify many previously unknown deployments. Our data collection phase includes (*1*) remotely testing more than 18,000 sensitive domains from ≈45,000 vantage points, yielding more than 374 million measurements and (*2*) adding publicly available censorship data, primarily 5 million measurements from OONI [25], a volunteer-run global censorship observation network affiliated with the Tor project that abides by ethical data collection norms. The diverse and complementary nature of these measurement methods and platforms allows us to paint a more complete picture of the global state of content-based filtering. We were able to detect 70 blockpage clusters that uniquely identify either the vendor that manufactures the filter or the actor that deploys it (if the former is unknown). We detect censorship filters in more than 100 countries—some at the national or ISP level, some on corporate or institutional networks—which restrict access to a range of content from pornography to political criticism. Among these 70 are blockpages from well-known commercial manufacturers including Palo Alto Networks, Cisco, and Fortinet, which contain an explicit indication of the vendor.

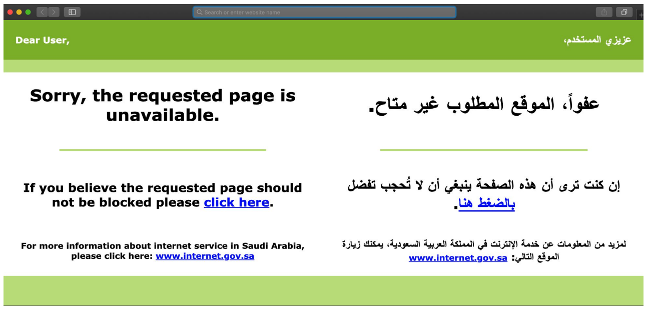
While FilterMap cannot discover every deployed censorship filter in existence—some may not be on visible networks, some may not inject blockpages, and some may evade our detection—it presents the most comprehensive view thus far of the deployment of filters used for censorship. Since filters are considered a dual-use technology, we do not make a determination on the appropriateness of their use nor the legal ramification of their deployment; for instance, we note that schools blocking access to porn and companies blocking access to Facebook is not comparable to governments blocking access to political and human rights content. Nevertheless, deciding on the appropriateness of the use of these filters is subjective and is outside of the scope of this paper. Regardless of the reason behind deployment of filtering products, the large-scale use of these technologies across many countries and institutions is alarming and emphasizes the need for regulators and the populace to gain visibility into the growth of content filtering technology.

An appealing application of our framework for the censorship measurement community (including projects such as OONI) is to use our analysis phase to reduce false positives. Our lightweight, semi-automated data processing methods can help identify measurement artifacts and noise that invariably appear in real-world networks. Moreover, designing effective circumvention tools requires considering both the capabilities of and the methods used by filters. Our system makes it easier for tool developers to create circumvention strategies based on empirical measurement by directing them to places where filters have been deployed and providing examples that trigger them. We hope this will help developers test and improve their circumvention strategies faster.

FilterMap's scalable and easily-deployable design enables longitudinal tracking of worldwide proliferation of filtering products. Interestingly, as shown in Section V-B, we identify 20 new blockpage clusters (in addition to the previous 70) in over 3 months of longitudinal measurements. We intend to maintain FilterMap as a source of longitudinal data for researchers monitoring global censorship. Our data and results are available at `https://censoredplanet.org/filtermap` and can be used by circumvention developers, advocacy organizations, and regulators seeking to understand and police the proliferation of censorship technologies.

II. BACKGROUND

While filters were originally adopted for purposes such as caching and security, they are increasingly being used by network operators to control their users' communication, most notably for censorship and surveillance. These products enable advanced filtering, particularly on application-layer data. In



Dear User,

عزيزي المستخدم،

Sorry, the requested page is unavailable.

عفواً، الموقع المطلوب غير متاح.

If you believe the requested page should not be blocked please click here.

إن كنت ترى أن هذه الصفحة ينبغي أن لا تُحجب تفضل بالضغط هنا.

For more information about internet service in Saudi Arabia, please click here: www.internet.gov.sa

لمزيد من المعلومات عن خدمة الإنترنت في المملكة العربية السعودية، يمكنك زيارة الموقع التالي: www.internet.gov.sa

Fig. 1: **Example of a blockpage:** Since blockpages often contain context about the vendor, authority, and/or reason behind blocking, they are more informative data sources than TCP/IP headers. ◇

contrast with IP blacklisting, which may result in collateral blocking, and DNS poisoning, which can be thwarted by using alternate DNS services, application-layer filtering is more accurate and effective, thus explaining its increasing prevalence. Filters can look into the content of HTTP requests and responses and disrupt the communication between client and server when undesirable keywords or domain names are observed. Advancements in filtering technologies have increased their capability to disrupt even encrypted traffic. For HTTPS, the Server Name Indication (SNI) extension to Transport Layer Security (TLS) leaks the domain name in plaintext. In case the traffic is fully encrypted (using ESNI [23], for instance), detecting pattern signatures of specific content such as Tor traffic could be used for filtering [30].

Filtering products, which are often manufactured by western companies, may come with subscriptions for updates, initialized blacklists, and a customizable blockpage—a notice that explains to the user why the content is unavailable—as shown in Figure 1. While acting on undesirable content, most filters provide information that may be used to identify either the vendor that manufactures the filter or the actor that deploys it. Researchers can then use different parts of the filter response such as the TCP/IP header, HTTP header, or the blockpage as a *signature* to investigate the filter's deployment.

Previous work on the Great Firewall of China identifies anomalies in TCP/IP header fields (e.g. IP ID and IP TTL fields) of injected packets as its signatures [60], [61]. Unfortunately, identifying filters from limited features of TCP/IP headers requires prior knowledge of existing vendor signatures in addition to cumbersome manual efforts. While we did collect TCP/IP data, we found blockpages, such as the one in Figure 1, more practical for extracting signatures and identifying vendors and actors. Blockpages are injected as a HTML response in the disrupted communication and often explicitly state the reason for blocking, possibly to dissuade users from trying to access the blocked content repeatedly.

Ethically collecting a widespread and diverse set of blockpages that result from triggering many globally deployed filters is challenging. One common method, which we refer to as direct measurement, is to seek assistance from on-the-ground and in-network volunteers [52], [31] who can run measurement software from their own devices. Tor's OONI project performs direct censorship tests from many volunteer users' machines to sensitive test domains. Another recent trend is to use remote measurement techniques that often use properties of existing protocols running on Internet infrastructure systems to measure Internet censorship [56], [46], [44], [43]. While direct and remote measurement approaches differ in depth, scale, and coverage, the data collected using global censorship measurement techniques are most likely to contain responses from many censorship filters, which we aim to detect.

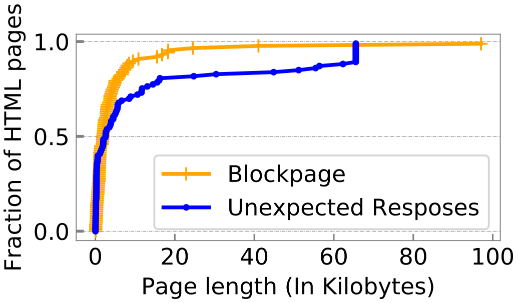


Fig. 2: **Page length in Kilobytes:** Blockpages and Unexpected Responses (e.g. common server errors) in Quack, Hyperquack and OONI data have comparable page lengths. ◇

As these filters act more commonly on HTTP(S) traffic, we were inspired by Quack, a recently introduced remote measurement system that efficiently detects application-layer blocking [56]. Quack uses servers running the Echo service on port 7, which echoes back any data sent to it. Quack uses this property and crafts HTTP-like requests in order to trigger a response from filters along a network path. Unfortunately, Quack does not generate legitimate HTTP(S) requests and it does not detect filters that only act on communications on port 80 or port 443. Hence, in addition to using Quack, we offer an upgrade for Quack where we use non-end-user-owned HTTP(S) web servers (in place of Echo servers) as vantage points and send genuine HTTP(S) requests with a variety of domain names exposed in the Host header or SNI extension. Normally, because these web servers are not expecting to receive requests for our test domain names, they respond with an error page (e.g. "Content Not Found"). However, if there is a filter in the path that is triggered by a test domain name, we might receive an injected blockpage. The blockpage is different from the typical server response, leading us to learn which domain triggered the filter. We name this new upgrade Hyperquack: Quack that uses Hypertext Transfer Protocol servers as vantage points (see Section III-A). Together, Quack and Hyperquack data contributed to 82 of the 90 blockpage clusters detected around the world by FilterMap.

Hyperquack upgrades Quack to detect filters that act on the HTTP(S) protocols, but it can only capture blockpages if the filter is configured to act on incoming (Measurement Machine → Vantage Point) requests. While the default behaviour of most filters is to act on all traffic regardless of the direction of the communication, there may be some that only act on outgoing traffic (Vantage Point → Measurement Machine). To detect these filters, we augment our measurement data with other public censorship datasets, specifically from the OONI project [25]. We use the OONI web connectivity test data, which detects censorship by comparing responses for a sensitive test domain with the response from a control measurement. Some of these responses contain blockpages

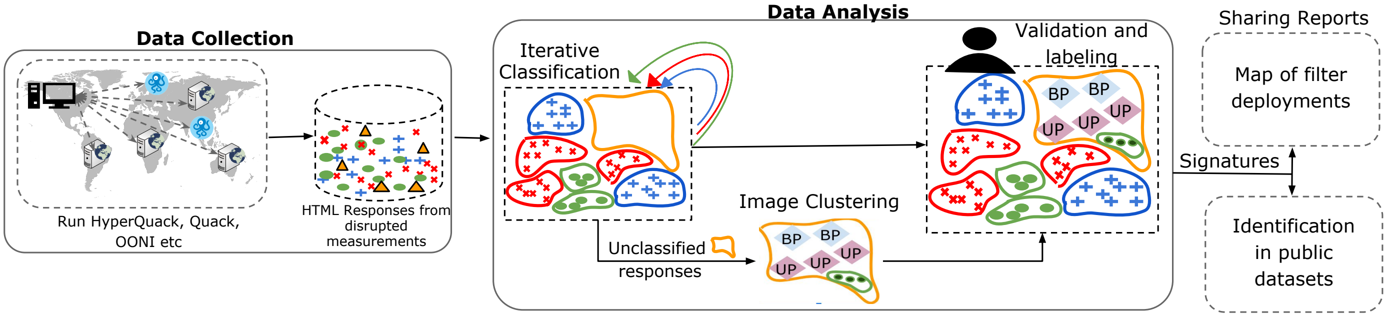


Fig. 3: **FilterMap Design**: FilterMap collects data using censorship measurement techniques. It then selects the HTML responses from disrupted measurements and runs iterative classification and image clustering to generate clusters of responses. We then manually label each cluster as a blockpage (BP) or an unexpected response (UP) and extract signatures. The outcome is a collection of signatures and the map of filter deployments. ◇

from filters. While OONI has sparse data, and an unclear number of volunteers that contribute to measurements—since their tests are anonymized to protect user privacy—we find 8 blockpage clusters unique to OONI data. Combining data from both direct and remote measurements helps us achieve the most comprehensive view yet on the deployment of filters used for censorship.

Another key challenge in identifying filters using blockpages is differentiating blockpages from *unexpected responses* such as server-side blocking errors (e.g. HTTP status code 403), page not found errors (e.g. status code 404), and DDoS checks from services hosted in CDNs [38], [53], [48]. Previous work [34] has used the metric of page length for identifying blockpages. However, later investigation on OONI data by Yadav et al. [62] found this metric to have high false positive rates due to natural variations in page length from dynamic and language specific content. They found that legitimate responses containing redirects were often misidentified as blockpages due to their short page lengths. Similarly, as shown in Figure 2, this metric is not suitable for the data collected in our work as typical server error pages are short (like blockpages) and have the same HTML structure. FilterMap's data analysis phase solves this problem by clustering pages based on content or visual similarity, rather than page length.

III. FilterMap Design

We introduce FilterMap, a framework that measures the deployment of filters that are configured to block traffic by responding with user-observable blockpages. Figure 3 presents a basic overview of the FilterMap's design, which includes the following key phases:

Data Collection Our data collection consists not only of a new measurement method, Hyperquack, but also Quack [56], and OONI's public censorship dataset [25]. Hyperquack and Quack run tests from our measurement location in North America to thousands of vantage points (i.e. HTTP, HTTPS, and Echo servers) in more than 190 countries. We complement the resulting dataset by downloading OONI's web connectivity test data collected by OONI volunteers. In Section III-A, we provide a detailed explanation of the data collection phase and Hyperquack's measurement technique.

Data Analysis To identify blockpages we use two techniques, Iterative Classification and Image Clustering, on HTML content extracted from disrupted measurements. These techniques help with generating large groups of pages that have either content similarity or visual similarity. Our experience suggests a filter often returns the same blockpage for all undesirable content. Therefore, we expect the injected responses to group in large clusters.

The clusters generated using both techniques are labeled by the signature that uniquely identifies the blockpages in that cluster. We generate these signatures manually, choosing a subset of the HTML page or Header that acts as a unique identifier. An example signature for the Barracuda filter is the presence of "<th>Barracuda NextGen Firewall:</th>" in the HTML body. Once the clusters are labeled, we detect the deployment of these filters based on our measurement data by looking at vantage points whose responses were injected with the corresponding blockpages. Additionally, these signatures can be used to search for injected responses in other datasets, although this may require further verification to prevent potential false positives. We describe one such experiment with public data from Censys [18] in Section V-D.

Note that the same filtering product might return different blockpages possibly due to different software versions or customizations. For instance, FilterMap generated 5 blockpage clusters that are associated with Fortinet products. In our results, we *aggregate clusters from the same vendor or actor* for clarity, which means Fortinet products would only count towards 1 of the 90 blockpage clusters.

Ethical Considerations Ethical considerations remain a major point of contention in censorship studies for both direct and remote censorship measurement systems [39], [11], [33], [56], [8]. FilterMap uses publicly available OONI censorship data and runs active measurements using remote techniques. Censorship measurement projects such as OONI that directly ask for on-the-ground collaborators to run measurements often seek informed consent from volunteers. Conveying the risk behind running sensitive tests require a great deal of human efforts to surmount volunteers' language and cultural barriers. The OONI project provides a summary of potential risks which they convey effectively to their volunteers and obtain informed consent. OONI's measurement data is publicly available and is used by many academic and non-academic research projects.

Remote censorship measurements such as Quack and Hyperquack use a fundamentally different approach that does not require volunteers' involvement. Considering the conceptual similarity of Quack and Hyperquack, we follow the ethical constraints described in VanderSloot et al. and only use *organizational* servers as our vantage points. We explain our vantage point selection process in detail in Section III-A.

While we cannot completely exclude the possibility that our organizational vantage points will be somehow penalized, we took several steps to reduce the potential risk. Because IRB considers remote measurement studies outside of their purview (as these studies did not involve human subjects or their personally identifiable data), we discussed the study's design with internal and external colleagues in our community. Like Quack, Hyperquack only establishes connections between our own measurement machine and organizational HTTP(S) servers. Because these administrators are likely to have more skills and resources to understand the traffic sent to their servers, the risk posed to them by these methods is lower than the risk posed to end users.

Moreover, we make it easy for anyone investigating our measurement machine's IP addresses to determine that our exchanged traffic is part of a measurement research experiment. We set up WHOIS records and a web page served from port 80, all indicating that the measurement machine was part of an Internet measurement research project based at our university. Over months of running measurements, we only received a handful of inquiries, and none indicated our probes caused technical or legal problems.

We also follow the best practices set forth by the ZMap measurement system [19] and limit the rate at which we conduct measurements using individual servers. To minimize the burden on servers we spread our measurements over many servers within a country, make a single request at a time, add delays between requests, and use a round-robin schedule to maximize the time between trials involving the same server. We use a fresh TCP connection for each request to minimize interference between requests. We also run our HTTP and HTTPS tests at different times so that servers used for both HTTP and HTTPS measurements do not receive simultaneous requests. On average, our probes caused servers behind filter deployments to trigger the filter 99 times a day, and a maximum of 240 times a day. As a point of comparison, two-thirds of Top Million sites cause requests to Google servers, and one-third to Facebook servers, so filters that block these large companies will be triggered more frequently by everyday browsing.

We conducted an additional check of running NMap on the Echo servers in the countries labeled as 'Not Free' by the Freedom House "Freedom on the Net Report" [26] and excluded servers whose labels left doubt as to whether they were infrastructural, increasing our confidence that we are not using residential vantage points in these countries.

A. Data Collection

The data collection phase aims to collect an extensive and diverse set of disrupted application-layer data—which most likely contains many blockpages from different filters. We achieved this by combining Quack, OONI and our new measurement system Hyperquack. While OONI data is public, we ran Quack following the exact measurement method and code described in [56] to collect data and detect disruption. The rest of this section describes the technical details about our new remote measurement technique, Hyperquack.

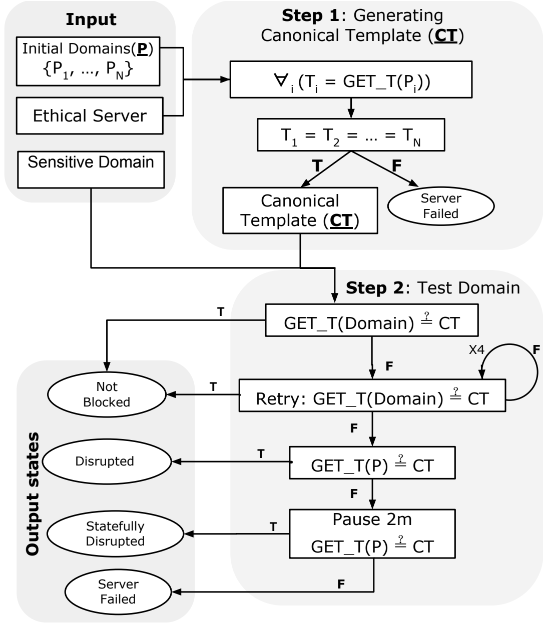


Fig. 4: **HyperQuack pipeline**: Given server, initial benign (sub)domains, and domain inputs, Hyperquack generates a canonical template `CT`, and then performs a set of trials, denoted `get_t`, to classify the test into possible output states. ⋄

Hyperquack Vantage Point Selection There are more than 50 million active HTTP(S) web servers around the world, with heterogeneous characteristics ranging from international CDNs to personal sites operated by individuals. We select servers from this pool with a focus on two properties: *Location Diversity* and *Ethical Soundness*.

Location Diversity: Our desire for location diversity is a property of the path as much as the remote vantage points. There are two important considerations on why the location of a server itself is not sufficient for our understanding of behavior. First, a subset of servers, including major CDNs, make use of anycasting. Anycasting describes the situation where an IP address is resolved to multiple physical hosts in different locations based upon the location of the client requesting content from the server [9]. This means that while the IP address may have a point of presence in a desired network, the connections from our measurement machine will not always be directed to that server. Second, servers may change their responses or behave differently based upon the location of a client. This means that we need to interact with servers from a single measurement machine to ensure that the behavior we observe is consistent across measurements.

Ethical Soundness: Aligned with the ethical considerations discussed in Section III and the ethical approach of [44], [56], we need to only use servers known to be *organizational*.

To address both constraints, we use the following approach for identifying organizational servers: Using the official list of Autonomous Systems, we enumerate optionally provided websites, which many providers include in their entries on peeringdb [45]. We choose these officially provided websites because as a primary point of contact, providers are incentivized to maintain these servers within their networks, rather than outsourcing them to cloud providers or maintaining them on their personal servers. Providers of this nature are established entities who most likely already have anomaly detection tools and abuse procedures as part of operations, meaning that they will be able to deal with our measurements. We also check one level of indirection in DNS from these providers. For instance, after identifying `level3carrier.com` as the official site for Level 3, we also check for a web server running on `mail.level3carrier.com`, and include that server in our list of potential vantage points. While this pruning step reduces the number of usable HTTP(S) servers from millions to thousands, it is ethically necessary for reducing risk. By selecting only web servers belonging to network operators from different parts of the world, we achieve both location diversity and ethical soundness.

Hyperquack Censorship Detection The Hyperquack measurement technique makes use of characteristics of servers that support the HTTP(S) protocol. The core concept behind this technique is that a web server's response can be predicted after sending some requests and observing the responses. For instance, when a client connects to the web front-end of `gov.uk`, requesting a domain that is not hosted there, the server responds with the message "unknown domain: Please check that this domain has been added to a service". Then, if the response for a sensitive domain diverges from the predicted response, it is indicative of the presence of a filter.

As illustrated in Figure 4, Hyperquack receives an input list of organizational web servers and a list of domains to test. Next, Hyperquack requests several bogus but benign domains, such as "www.test<rand int>.<rand tld>" and "www.example-<randint>.com", to create a *template* for the expected response. Server response may vary because HTTP(S) responses typically contain dynamic elements such as cookies or timestamps; Therefore a template of server's predictable behavior is generated by removing these common dynamic elements and occurrences of the test domain. For HTTPS servers, the templates include certificate and cipher suite choice. If the templates for all of these benign domains match, Hyperquack classifies the web server as consistent and saves the *canonical template*. We were able to generate canonical templates for 87% of the organizational servers. Server behavior may also vary based on the requested subdomain. The most common example is when the web server adds or removes the "www" portion of requested domains on its error pages. Hyperquack uses a variety of control subdomain patterns when testing for server consistency. We make requests of the form "<subdomain>.example<rand>.com" for all subdomains in our test list, and compare these replies with the generated canonical templates. The rates of mismatches are below 1% for all subdomains except "webmail.sso" and "mail". Fortunately, only a negligible number of test domains—precisely 18 out of 18,736—have these subdomains.

The measurements start once the canonical templates have been generated. Hyperquack tests each domain for interference across the consistent organizational servers (Step 2 in Figure 4). It begins each trial by making a GET request for the domain and generating the reply template. If it matches the server's canonical template, the trial is done. In this case, the output state is *Not-Disrupted*. If the reply does not match, it makes up to four retries to accommodate temporary network changes. Because network interference is sparse and organizational servers are generally reliable, out of all the *Not-Disrupted* cases 99.5% of the trials end after a single request, allowing the system to scale.

If all retries fail to match the canonical template, Hyperquack checks to see if the server is still behaving consistently by requesting a randomly generated control domain. If the reply for the control domain matches, the server is still consistent and Hyperquack attributes the disruption to the test domain. The output state in this case is *Disrupted*, indicating either the presence of a filter or an unexpected response for the domain. We explain how our analysis identifies and separates filter responses and unexpected responses in Section III-B. In our measurements, we observed filters practicing various types of disruptions, most commonly injecting a blockpage, injecting a TCP RST, or forcing the connection to timeout. In HTTP tests, more than 50% of filters' responses contained a blockpage. If none of the responses match the template, the server is deemed to be behaving inconsistently and Hyperquack does not use the server for future trials. In this case, it marks the output state as *Server Error*. We observed on average a failure rate of only 1.5% for HTTP and 1.9% for HTTPS after completion of full measurements. Because the matching criteria for HTTPS templates is stricter (it contains more features), the higher failure rate is expected.

Following redirects Some filters are configured to respond with a redirect to their blockpage. Therefore, we anticipate observing some web redirects in our collected data. In the post-processing step, we follow web redirects and add the HTML retrieved from these redirect URLs for the analysis phase.

B. Data Analysis

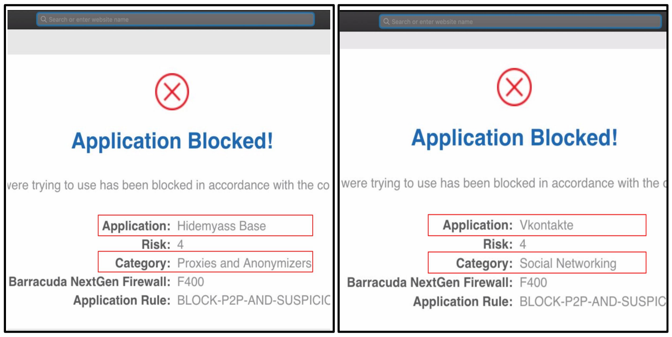
In this phase, FilterMap generates clusters for blockpages using iterative classification and image clustering, thus significantly reducing the manual effort required in identifying and labeling blockpages. The iterative classification prioritizes classifying pages that occur frequently and image clustering is highly effective in grouping HTML pages with dynamic content like embedded advertisements which have considerably different HTML code but render visually similar pages.

We use all the HTML responses marked as *disrupted* from all the collected datasets. The HTML content can either signify a blockpage or an unexpected response. Unexpected responses are anomalous responses—such as server-side blocking errors (e.g. status code 403), page not found errors (e.g. status code 404) and DDoS checks from services hosted in CDNs—that occur due to unexpected vantage point behavior towards some domains. Our manual labeling confirms these responses do not originate from filters. For example, in Hyperquack we find instances where a web server hosts one of the test domains and thus sends back the genuine content of the domain that is different from our canonical template.

Iterative Classification Considering the enormous amount of collected data, any clustering approach would take

Algorithm 1 : Iterative Clustering

```
block-patterns ←[ ]
ur-patterns ←[ ]
unclassified ←[ ]
procedure ITER-CLASS(param,HTML-pages,block-patterns,ur-patterns)
    foreach HTML in HTML-pages
        if get_pattern(HTML) ∈ {"Blockpage", "Unexpected Response"} then
            next
        Unclassified ← HTML
    unclassified_length = $len$(unclassified)
    assert(param $> 0$)
    groups = make-group(unclassified)
    unclassified ←[ ]
    flag = False
    foreach group in groups
        if $len$(group)/$len$(unclassified_length) $* 100 >= param$ then
            flag = True
            pattern,type ← get_pattern(group),label_manually(group)
            if type $==$ "$Blockpage$" then
                block-patterns ← pattern
            else if type $==$ "$UnexpectedResponse$" then
                ur-patterns ← pattern
        else
            unclassified ← group
    if flag $==$ False then
        return unclassified
    else
        return ITER-CLASS(param,unclassified,block-patterns,ur-patterns)
```


Search or enter website name

Application Blocked!

vere trying to use has been blocked in accordance with the co

Application: Hidemyass Base
Risk: 4
Category: Proxies and Anonymizers
Barracuda NextGen Firewall: F400
Application Rule: BLOCK-P2P-AND-SUSPICIC

Search or enter website name

Application Blocked!

were trying to use has been blocked in accordance with the c

Application: Vkontakte
Risk: 4
Category: Social Networking
Barracuda NextGen Firewall: F400
Application Rule: BLOCK-P2P-AND-SUSPIC

Fig. 5: **Dynamic elements in blockpages**: Zoomed blockpages injected for hidemyass.com and vkontakte.ru. Image clustering uses the visual similarity to correctly group these pages together. ◇

considerable computing time and resources to cluster blockpages. Inspired by the iterative fingerprinting methodology described in Weaver et al. [59], our first systematic step is to reduce the number of elements that need to be clustered simply by grouping recurring HTML pages (which contain both blockpages and unexpected responses). As shown in Algorithm 1, the iterative classification algorithm starts with an empty *blockpage set*, an empty *unexpected responses set*, and an *unclassified set* initially containing all the HTML responses.

The first run results in different groups, where each group represents a set of matching HTML pages, and often picks up a handful of widely deployed filters. Since data from the data collection phase may contain over 500,000 blockpages, the first run can generate upwards of thousands of groups. In each iteration, to find candidate groups for manual labeling, we look at the group sizes above a certain empirical threshold, i.e. groups with size above a certain percentage of the *unclassified set* are candidates for labeling. In this study, we find that a threshold of 10% of the size of the unclassified set gives an acceptable coverage rate with a very low number of iterations. The labeled groups (blockpages or unexpected responses) are added to the sets accordingly and generated labels (signatures) for the groups are stored. The iterative classification algorithm recursively continues on the remaining unlabeled instances. The iteration stops when no remaining groups are big enough to pass the threshold.

Image Clustering As mentioned previously, iterative classification fails to classify responses that occur infrequently or have dynamic content (e.g. Figure 5). To minimize human involvement in classifying these instances, we use a deep learning model to extract feature representations, or embeddings, from screenshots taken from rendered responses, and cluster these screenshots using the resulting embeddings. Previous work [7] rely on neural network models (e.g. an autoencoder) to extract compact representations of a webpage's content. We adopt a similar approach by using a pre-trained image recognition model, more specifically ResNet50 convolutional neural network [29] trained on the ImageNet [16] dataset, to obtain embeddings of length 2048 from each screenshot. We extract embeddings from the layer before the softmax layer, which is used to output predictions for ImageNet classes. This step is intended to extract meaningful representations from the screenshots, while reducing the dimensionality of the original images, which in turn reduces computational requirements for the clustering algorithm. We employ image clustering as opposed to clustering using document similarity, because we found instances of the same blockpage returned in different languages (based on configuration), and because we expect a high visual similarity between blockpages from the same filter.

Next, we use the DBSCAN [24] algorithm to cluster the reduced feature vectors obtained from all the screenshots, using a minimum of five samples to form a cluster. Clusters are extracted by using five different ϵ values, which represent the maximum euclidean distance of embeddings of two samples to be considered as in the same neighborhood. The reason we chose to use different ϵ values (10, 5, 2, 1, and 0.5) for the clustering process is that a relatively low ϵ would result in separating content-heavy pages that should otherwise be clustered together, while using a relatively high ϵ would group together all the pages with the same background color and a small amount of content. It is also worth noting that a cluster obtained using a lower ϵ is always going to be a sub-cluster of one obtained using a higher ϵ, allowing us to inspect each cluster and its corresponding sub-clusters, and choose an appropriate ϵ level accordingly. The clusters are then labeled using the signatures generated based on what we gather from the HTML content of the blockpage. All the clusters that do not have any indication of being a blockpage are labeled as unexpected responses.

With iterative classification and image clustering, we tremendously reduce the effort needed to identify blockpages. The only manual effort is in labeling the groups with the appropriate signature. This required looking at one HTML page from approximately 200 groups. Fortunately, enriching these known blockpage sets over time can speed up the data analysis phase by pre-processing the next generation of collected data using regular expression matching of their known signatures.

IV. Experiment Setup and Evaluation

In this study, we performed two large-scale measurements; the first is a latitudinal measurement that expands over several

	HTTP	HTTPS	Quack
Initial Set	9223	6200	36000
Experiment Set	9063	6070	33602
Number of Countries	215	204	75
Median / Country	11	13	151
Number of AS	4558	3442	3463

TABLE I: **Characterization of Vantage Points:** The experiment set consists of servers that passed consistency tests for all domains without failing. The reported country and AS is based on the qualified servers. ◇

	BR (%, #)	UR (%, #)	UC (%)	# of Iterations
HTTP	(56.51%, 27)	(39.39%, 105)	4.10%	3
HTTPS	(3.48%, 5)	(83.83%, 67)	12.70%	1
Quack	(93.08%, 34)	(4.8%, 116)	2.12%	2
OONI	(13.02%, 16)	(43.27%, 44)	43.71%	2

TABLE II: **Iterative Classification Evaluation**: This table shows the (percentage of responses, number of groups) of HTML responses that were classified as blockpages (BP), unexpected responses (UR) or were unclassified (UC) in the three-week measurement data. ◇

thousand sensitive test domains with the goal of triggering many filters; the second, longitudinal measurement, expands over several months, hence displaying the effectiveness, scalability, and sustainability of FilterMap.

Latitudinal Measurement We initially performed a three-week measurement in October 2018 for Hyperquack and Quack. We use entries from the Citizen Lab Test List (CLTL) [10] as our potentially-blocked test domains. CLTL is a curated list of websites that have either previously been reported unavailable or are of interest from a political or human rights perspective. CLTL includes 102 country-specific lists, categorized into 33 categories from public health to gambling. We used the list as on October 4, 2018, which contained 18,736 unique domains. We combine this data with the OONI web connectivity test data from the same period. OONI's volunteers performed connectivity tests to 15,828 unique domains during that time.

Longitudinal Measurement We ran Hyperquack and Quack scans twice a week from November 2018 to January 2019 to show FilterMap's ability to continuously detect filters over an extended period. We tested ≈2100 domain from the CLTL global list which contains domains of interest across many countries, and the top 1000 most popular domains from the Alexa Top 1M Dataset [3]. We used the Alexa list based on its applicability to user-centric measurement studies and to detect blocking of popular domains which may not be added to the CLTL immediately. Considering the continuous nature of these scans, we reduced the number of test domains to prevent unnecessary load on vantage points.

Vantage Points To find HTTP(S) servers we followed the process outlined in Section III-A. For a representative test run in October 2018, this process yielded 11,700 organizational web servers. Not all of these servers behaved consistently enough to characterize their behavior. After removing CDNs and IPs blacklisted due to abuse complaints, we were left with 10,470 servers. Consistency checks further reduced the number of vantage points to 9,223 servers for HTTP measurements and 6,200 servers for HTTPS measurements. From Table I, we observe that 98% of the servers complete all measurements without failing consistency checks. For Quack, we found over 45,000 echo servers using custom ZMap [19] scans. Of these, we excluded 92 servers that we could not confirm are organizational according to our ethical constraints (Section III). Our initial vantage point set consisted of 36,000 echo servers which remained stable during the consistency checks.

Data Collection Evaluation Our measurement machine performs measurements for 1,000 domains on all HTTP(S) and Echo servers in under 10 hours— While we can speed up by sending many requests in parallel, we abide by best practices and limit the rate of sending requests (as described in III). We find that 95% of the measurements complete within 3 and a half hours. The remaining 5% of measurements take up to 6 hours more due to slow server responses.

Data Analysis Evaluation Clustering approaches often take considerable computing time and resources and our analysis phase had to process large quantities of data. With multiple strategies to optimize performance as described in Section III-B, each iteration of the iterative clustering algorithm only runs for approximately 25 minutes on a dataset of 1 million HTML pages. Through manual trials, we find that a threshold of 10% of the size of the unclassified set gives an acceptable coverage rate with a very low number of iterations. Detailed performance analysis per data collection tool is presented in Table II. The high percentage of unclassified responses in OONI is due to difference in data collection method, as OONI's volunteers test different domains. FilterMap identified a total of 82 blockpages using iterative classification.

For image clustering, note that the time complexity of the DBSCAN algorithm for high-dimensional data is $\mathcal{O}(n^2)$, where n is the number of samples being clustered. For our largest set containing ≈80,000 screenshots, the clustering algorithm takes 5 minutes per ϵ level (25 minutes for all levels). To evaluate the accuracy of image clustering, we manually inspect the corresponding clusters for 100 randomly picked screenshots and examine whether they have been correctly grouped with similar images for any of the utilized ϵ values. For HTTP (HTTPS) responses, we observe that 84 (90) samples were correctly grouped with similar screenshots. Interestingly, we found no false positives (i.e. a sample that looks different from the rest of the screenshots) in the respective clusters for these samples. The remaining 16 (10) samples corresponded to unexpected response pages with a white background and, optionally, a single short line of content (e.g. "Bad Request", or "Page moved permanently"), and were grouped together in one cluster (two clusters) for the lowest ϵ value. We further inspected all clusters obtained from Quack and OONI measurements, and similarly observed that the only case where our technique did not group similar images was for screenshots with a white background and a very short line of content. Inspecting all HTML pages inside blank page clusters confirms they are not blockpages, hence the imprecision of our approach on blank pages does not affect our ability in discovering blockpages. Upon inspection of the generated clusters from image clustering, we additionally identify 8 new blockpages.

V. Results

In FilterMap's data collection phase, we tested more than 18,000 sensitive domains from ≈45,000 vantage points, yielding more than 374 million measurements, which is augmented with 5 million measurements from OONI. FilterMap's data analysis phase generated 90 blockpage clusters in which the blockpages identify either the commercialized vendor that manufactures the filter or the actor that deploys it (e.g. government, ISP, or organization). These filters are located in many locations in 103 countries revealing the diverse and widespread use of content filtering technologies.

The signatures generated for 87 of these blockpages were previously unknown. FilterMap extracted 38, 49, and 21 blockpages from Hyperquack, Quack, and OONI data respectively. Hyperquack triggered a large number of commercial filters—products with an explicit indicator of the vendor in its blockpage—being used for censorship. This is because the default configuration of most of these commercial filters acts on HTTP(S) traffic on port 80 and 443. Hyperquack data detected deployments in three times as many countries as Quack and OONI. Because of the large number of Quack vantage points in ISP networks, Quack was helpful in detecting a large number of filters deployed in ISPs, especially in Russia (38 out of the 49 filters detected by Quack were deployed in Russian ISPs).

We observed blockpages in 14 languages: filters which are locally configured return blockpages in regional languages, while most exported from western countries have a standard blockpage in English. Some blockpages have content in multiple languages, e.g. local content and advertisement added to the template blockpage in English. We also observed the same blockpage being returned in different languages, suggesting vendors customize template blockpages for their customers. Fortinet products have the largest presence in many countries in our data, with blockpages observed in both English and Chinese. Note that image clustering was able to group these instances into the same cluster, which would not have been possible with text-based clustering methods.

We manually categorized these blockpages as Legal (containing a notice about blocking based on a law or court order), Informative (not Legal, but containing information regarding blocking), or Unclear (not indicating that the content is being blocked)—to prevent potential bias, three authors independently categorized the blockpages and arrived on a consensus. The breakdown of blockpages is as follows: 60% Legal, 32% Informative, and 8% Unclear. The large number of blockpages with explicit reference to law or court order suggests most networks that follow orders of authorities are eager to inform their users of the reason behind blocking. This is especially striking in Russia where all the blockpages from 41 Russian ISPs cited Federal Law as the reason for blocking. We observed that most filters deployed in organizations fall under the informative category. Finally, 8% of Unclear blockpages either show an error page or a blank page instead of an indication of blocking. Upon looking at the raw HTML content, we see some indication of blocking that is not easily visible to the user. We observed 48 filters returning status code "302 Found" which is a common way of redirecting requests to a blockpage. Apart from redirects, the most common blockpage status codes were "200 OK" (15) and "403 Forbidden" (14).

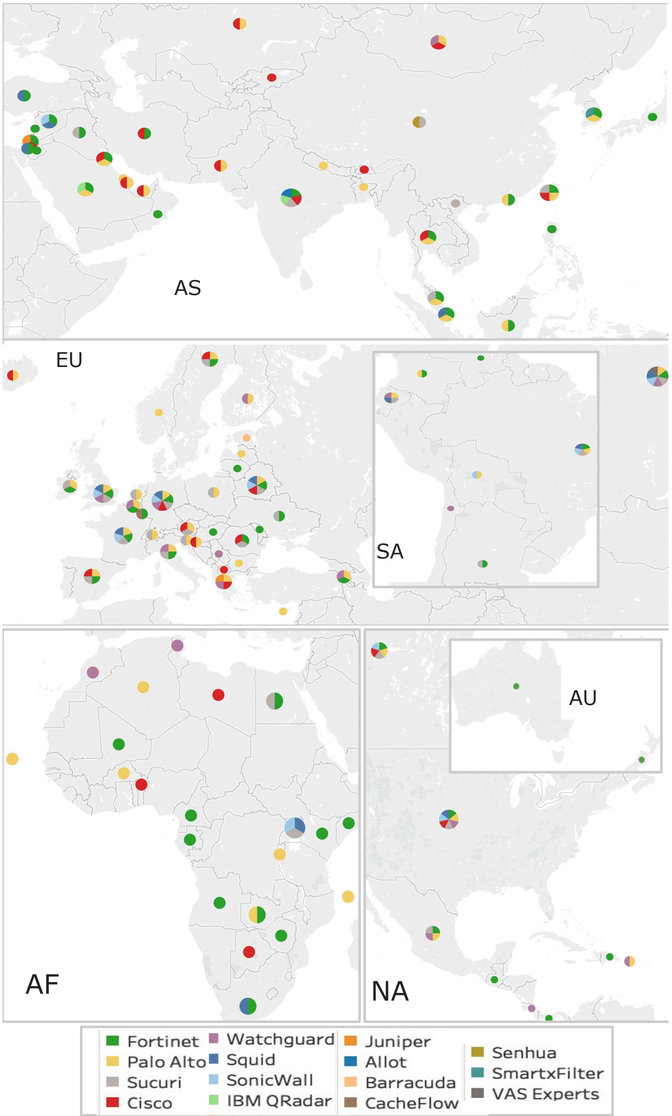


Fig. 6: **World map of commercial filter deployment:** Circle size represents number of products detected in that country. ◇

A. Filter Types

Blockpages (and the corresponding signatures) can help identify the vendor that manufactures the filter, or the actor that deploys it. *Commercial filters* are well-known content filtering technologies that are available on the market and have trademarks such as "<TITLE>Juniper Web Filtering</TITLE>" or "<h2>Powered By FortiGuard</h2>". These products are mostly deployed by actors such as organizations, ISPs, and governments. *Filters with Government Blockpages* have a presence on a nation-wide scale (e.g. Bahrain). *Filters with ISP Blockpages* are deployed by ISPs for restricting access to certain websites for their users. ISPs may block content to comply with directions from telecommunications authorities or censorship laws (e.g. Russia, India). *Filters with Organizational Blockpages* have information about the organization that

Country of Origin	Commercial filter
Israel	Allot
China	Senhua
Republic of Korea	SmartxFilter
Russia	VAS Experts
United States	Barracuda, CacheFlow, Cisco, Fortinet, IBM QRadar, Juniper, Palo Alto, SonicWall, Squid, Sucuri, WatchGuard

TABLE III: **Location of commercial filters' headquarters** ◇

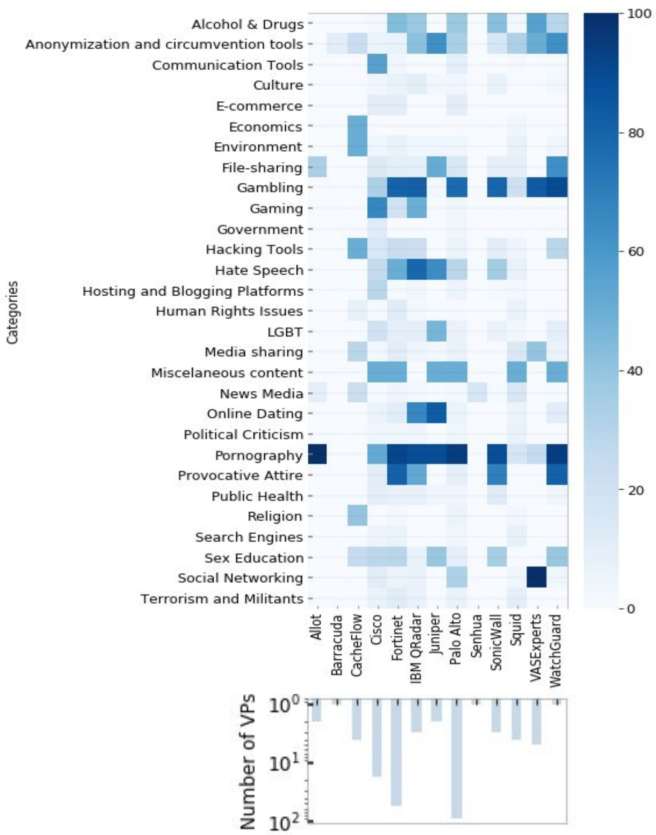


Fig. 7: **Categories blocked using commercial filters:** The heatmap shows the median blocking percentage of domains in the global test list by vantage points behind each commercial filter. The bar chart shows the number of vantage points (in log scale) contributing to the median value. Categories such as Pornography, Gambling, and Anonymization tools have high blocking rates, showing that these products are increasingly used for censorship. ◇

deploys or owns the filter, such as a university. Finally, there are some filters where the vendor or actor is *Unknown*.

1) Commercial filters: FilterMap identified 15 commercial filters in 102 countries, as shown in Figure 6. The availability and ease of deployment of commercial filters has galvanized the process of restriction of Internet freedom in many countries. About three quarters of these products explicitly state the vendor name making them easy to identify, while the others required further investigation such as following links in the blockpage and looking at the alternate text behind logos. Many countries have no export or import restrictions on these products, allowing for the free flow of this technology to other countries. The most popular product, Fortinet, has a presence in at least 60 countries.

While a small portion of commercial filters are locally manufactured and deployed (e.g. VAS Experts in Russia), most suppliers are headquartered in the United States, as shown in Table III—A company's headquarter location is where the company has legal responsibility for decisions made. The export of these products to countries with poor records in Internet freedom is a cause for concern. 36 out of 48 countries labelled as 'Not Free' or 'Partly Free' by the Freedom House "Freedom on the Net" report [26] appear in Figure 6. The unrestricted transfer of content filtering technology has led to high availability, low cost, and highly effective filtering techniques becoming easier to deploy and harder to circumvent.

As shown in Figure 6, some commercial filters are more popular in specific regions. For instance, FilterMap detected a high presence of Fortinet and Cisco products in African countries where Internet Freedom has been a cause for concern [20]. Similarly, Palo Alto products are deployed in a large number of countries in Europe, where the Internet is generally thought of as free.

To characterize the kinds of content being blocked by commercial filters, we extracted the categories of domains in the Citizen Lab Global Test List [10], and computed the blocking percentage for each category per vantage point in Hyperquack and Quack (OONI volunteers rarely test the CLTL lists—running tests for 18,000 domains can become costly, especially in developing countries. The median value of this blocking percentage per vantage point is shown in the heatmap in Figure 7 alongside the number of vantage points behind each filter. Pornography and Gambling websites are most commonly blocked across all products, followed by websites featuring provocative attire and anonymization tools. Interestingly, we observed that the Russian product VASExperts is being used to block access to many social networking websites, including popular ones such as LinkedIn.

2) Filters with Government Blockpages: We identified primary blockpages in Bahrain, Iran, Saudi Arabia and Republic of Korea that previous reports have established as government specific blockpages [14], [4], [65], [1]. Refer to Appendix B for the observed blockpages. In Hyperquack data 97.1% of the disruptions in Iran were caused by the Iran firewall, while the Bahrain and Saudi Arabia firewalls contributed to 71.2% and 80.2% of the disruptions, respectively. Note that China, a well-known censoring country, does not appear here because we observed in agreement with previous work [22] that the Great Firewall resets connections instead of responding with a blockpage. We also do not detect a few other known firewalls such as the ones in Qatar and Turkey, because they either disrupt traffic using techniques other than application-layer blocking (such as DNS tampering or TCP/IP blocking), or because their blockpages do not explicitly indicate governments' involvement. Pornography websites, gambling websites, and anonymization and circumvention tools are three of the most common categories of domains disrupted by filters in these countries.

The level of detail in the blockpage returned by the filter in Saudi Arabia (shown in Figure 1) is surprising, but shows the desire of the authority to provide explicit information regarding the connection disruption, possibly to dissuade users from trying to access blocked content again. Although the Bahrain blockpage does not give any information about the vendor,

Countries	ISP with filter Deployment
Ivory Coast	MTN
Iceland	STEF
Bahrain	VIVA
Mauritius	Airtel
India	Court Order India
Kyrgyzstan	Elcat
Saudi Arabia	STCS
Yemen	TeleYemen
Kuwait	Zain ISP
Colombia	ERT
Pakistan	Wi-tribe
Republic of Korea	SKT
United Arab Emirates	Etisalat
Belgium	Telenet
Russia	Convex, RSVO, Piter-Telekom, Wiland, Kamenkstel, Avantel, Orion Net, Sivash, Strela Telecom, Infolink, Intertax, East Media, Sky@Net, Sevstar, Altegrosky, DTEL, Goodline, Dianet, Maglan, Skynet, Sibitex, Novotelecom, Yota, DSI, Kristel, ITNet, Westlan, UGMK-Telecom, Spacenet, ACME, Iterika, Mosnet, Metroset, Redcom, Bashtel, TSCrimea, IKS, Divo, Beeline, MTS, Flex

TABLE IV: **Filters deployed in ISPs** ◇

Organization	Country
Brazilian Federal District Government	Brazil
Sun TV Network	India
AUIS	Iraq
Gyeongsang University	Republic of Korea
Elko	Latvia
National University Singapore	Singapore
Northwestern University	United States
Uniminuto	Colombia
Pustekkom	Indonesia
Itgrad	Russia

TABLE V: **Organizational filter deployments** ◇

reports by Citizen Lab [49] indicate that the product used by Bahrain is manufactured by Canadian vendor Netsweeper [14]. Recent reports from South Korea have indicated a rise in SNI-based censorship and a shift away from DNS-based censorship, following the trend in the increase of keyword-based blocking [6].

3) Filters with ISP Blockpages: These filters are deployed by ISPs to restrict access to users of the ISP network. ISPs play an important role in Internet censorship as many countries practice decentralized Internet control at the ISP level. For example, ISPs in India block content under guidelines from the Department of Telecommunications [62]. ISP blockpages most often contain legal information to describe the reason behind blocking content. The blockpages are also predominantly in the local language. FilterMap was able to detect 41 ISPs restricting access to content in Russia, as shown in Table IV. Other than Russia, we found filters deployed in popular ISPs in a large number of countries in Asia.

4) Filters with Organizational blockpages: Some blockpages do not contain any indication about the vendor nor about government or ISP policy but do contain content indicating that the filter has been deployed as a result of organizational policy. Although these filters do not restrict content at a nation-wide scale, detecting them can bring transparency to corporate and social organizations' policy, and can aid in highlighting cases of egregious blocking of content. FilterMap was able to detect the presence of 10 filtering deployments of this kind, as shown in Table V. Six of these deployments were in universities.

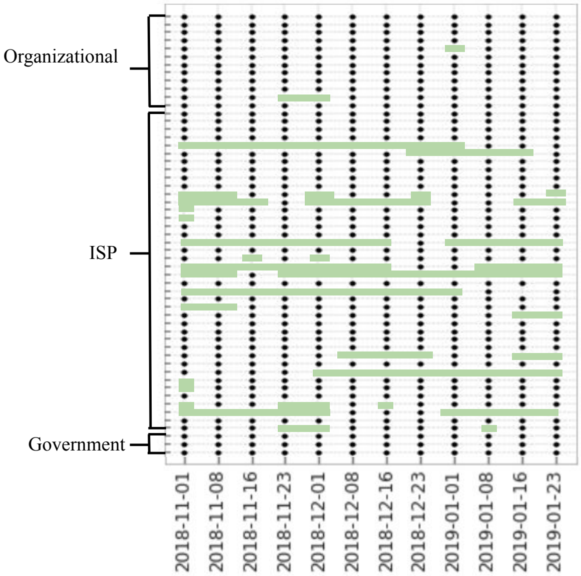


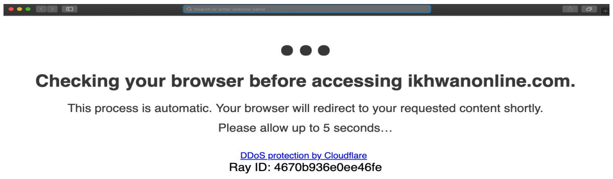
Fig. 8: **Filter deployments detected over time:** The scan dates are aggregated to week-level. Each dot represents the presence (a green line indicates an absence) of a blockpage for the corresponding category. ◇

5) Unknown filters: FilterMap detected the presence of 6 blockpages that do not contain any identifiable information. For example, one of the the HTML responses in this category only contains a title "Warning" and a line "This webpage has been blocked".

B. Longitudinal data

FilterMap is a scalable, lightweight, and easily-deployable tool that can detect filter deployments over time. During our 3 month longitudinal measurement, we collected data using Hyperquack and Quack semi-weekly. We discovered 20 additional blockpage clusters apart from the ones discovered from our latitudinal measurements.

The manual effort required to label clusters reduces over time as more blockpages and unexpected responses are added to our database of known regexes. As expected we observe very few new large clusters due to the deployment of new censorship systems. Most commercial filters (except Allot) were detected in every scan. Allot was only detected using one vantage point, and some scans did not include that vantage point due to churn in server selection. For filters with organizational, ISP, and government blockpages, Figure 8 shows the presence (or absence) of each blockpage in each run over 3 months. We did not observe any Saudi Arabia blockpages in our longitudinal scans because of lack of vantage points, but we detected the other Government blockpages in all scans. ISP and organizational blockpages have infrequent absences that can be explained by churn in our vantage point selection sources, routing changes in the path, and changes in policy—as filtering decisions change over time. The new deployments discovered were because of the inclusion of new vantage points, or due to configurations being modified to respond with a blockpage. This not only shows the capability



Checking your browser before accessing ikhwanonline.com.

This process is automatic. Your browser will redirect to your requested content shortly.

Please allow up to 5 seconds…

DDoS protection by Cloudflare

Ray ID: 4670b936e0ee46fe

Fig. 9: **An example of an unexpected response** ◇

Filter	**# of IPs**	**# of countries**
Barracuda	29	4
Fortinet	10,748	151
Juniper	41	2
Palo Alto	3,087	72
Watchguard	211	28
Cisco	1,434	63
IBM QRadar	22	5
SmartxFilter	33,639	2
Sucuri	24	8
Squid	1	1

TABLE VI: **Signature matches in Censys**: This table shows the number of IPs in Censys whose response matched with one of our commercial filter signatures. ◇

of our technique, but also highlights the need for continuous detection of filters for monitoring their proliferation.

C. Unexpected Responses

In disrupted responses from Hyperquack, Quack, and OONI, FilterMap detected unexpected responses in addition to blockpages. Figure 9 shows an example of an unexpected response detected in all of our datasets. This page contains an explicit note indicating that the site is hosted on Cloudflare and protected against DDoS attacks, and this does not indicate any kind of censorship. Since this response is different from the control response, Hyperquack, Quack, and OONI label it as disrupted. There are many different types of unexpected errors indicating server-side blocking (status code 403), "Not Found" errors (status code 404), DDoS checks, etc. One appealing application of our lightweight data processing method is to help distinguish these errors from censorship; A tool that can be easily adopted by the censorship measurement community (including projects such as OONI and Quack) to identify measurement artifacts and noise from their censorship measurement data.

D. Using signatures in other public datasets

In this work, our goal is not to produce compact signatures of blockpages that can be applied to other datasets, but rather to detect deployments of filters around the globe by identifying unique blockpages. While the signatures generated by FilterMap for identifying blockpages can be used in datasets other than the ones described thus far, the discovery might not indicate the use of these filters for censorship. To demonstrate this potential, we performed an experiment by searching for these signatures in Censys data [18]. Censys collects HTTP(S) responses obtained from every public IP address on the Internet that hosts a web server. We downloaded Censys HTTP and HTTPS measurement data on September 12, 2019 and searched for our signatures in the responses recorded by Censys. Out of 100 signatures (corresponding to 90 blockpages), 19 unique signatures (corresponding to 14 blockpages) matched with at least one Censys measurement response. The number of IP addresses that returned a response matching signatures for the commercial filters is shown in Table VI. We find a large number of blockpages from Fortinet and Palo Alto in Censys data. Filters were found in 154 countries probed by Censys, many of which were not discovered using Hyperquack, Quack and OONI data. Since Censys probes every single IP address, it contains a more complete view of the Internet. However, Censys does not request commonly-censored content and does not measure for censorship, thus detecting many firewalls performing access control and DDoS protection.

Using signatures to detect blockpages in public datasets such as Censys may result in false positives. To obtain a rough estimate of the number of false positives, we selected a sample of Censys responses matching our signatures and manually parsed each response. Specifically, we utilized disproportionate stratified sampling, randomly selecting up to 10 responses for manual checking from each set of matching responses for a signature. 154 responses were selected in this manner, and after manual verification, none were determined to be false positives. However, we discovered some blockpages that have slightly different content and also found blockpages in new languages, such as a Spanish Fortinet blockpage.

We observe corroboration in results between our measurements and Censys data. For instance, most IP addresses that return a SmartxFilter blockpage are in South Korea (where it is manufactured) as also detected by Hyperquack data. However, Censys data also shows a few IP addresses in Iran returning the SmartxFilter blockpage, indicating that the product may be exported to Iran. Our experiment with Censys shows that signatures generated by FilterMap can be used in other public datasets for identifying blockpages. The experiment also shows large networks in almost all countries in the world employing the use of filters, providing more insight into the proliferation of these systems.

VI. Related Work

While we extensively discussed related work that directly affected our design choices in Section II, we discuss other relevant works in this section. Many measurement systems utilize lists of keywords for testing censorship. On the web, domain names are commonly used as a proxy for services, and are typically drawn either from lists of popular global domains [3], or from curated lists of potentially sensitive domains [10]. To conduct measurements on a sufficiently large corpus, and to maximize comparability, our system uses both of these sources.

Detection of keywords more broadly has made use of corpora extracted from observed content deletion, along with NLP and active probing. [64], [27], [13]. Previous systems determining such keywords have largely focused on individual countries and services, especially related to Chinese social media such as Weibo and TOM-Skype [35], [12].

Deep packet inspection and application-layer disruption have become standard practice online [17]. Asghari et al. [5] find support for their hypothesis that nations pursing censorship are likely to push deployment of application-layer filtering technology. ONI reports have shown that filters manufactured by western companies are used by countries in the Middle East for effecting censorship [40]. OONI reports on application-layer censorship in 12 countries with identified vendors [50], and the Tor project has noted keyword-based blocking in at least 6 countries [2]. Prior work by Marczak et al. [36] involved acquiring and analyzing a Sandvine/Procera PacketLogic product in a lab. Although they had success detecting the deployment of that product, their process is cumbersome, and it will only identify known filters. A recent report traces the diffusion of the Chinese and Russian models of information control to 110 countries [55], mainly through manually identifying a chosen set of middleboxes that are exported from China and Russia using public network measurement data.

Techniques similar to the one described in this paper have been used in a limited context for identifying specific products and web proxies [37], [59], or to find where products are deployed within a country or network [14], [1]. Dalek et al. [15] explored an alternative approach to investigating known vendors by first looking at known blockpages and manually extracting signatures. They were then able to look at publicly available Internet-wide scanning data from Shodan [47] and the 2012 Internet Census [32] to discover deployments. Although their results show the deployment of four vendors across multiple countries, their approach required a manual validation process, requiring in-country testing to confirm their findings.

VII. Discussion and Conclusion

In this paper we present FilterMap, a framework for identifying and monitoring filters based on the blockpages they display.

Limitations Though FilterMap finds a significant number of new filter deployments, it cannot discover all filters deployed around the globe. Some network paths are not covered by our measurements to the selected vantage points, some deployments are not triggered by our data collection techniques, and some manufacturers or actors cannot be identified through the injected blockpages. Ethical considerations limited the selection of our HTTP(S) vantage points to the ones we were confident are not end-user-owned. This selection drastically reduced the number of Hyperquack vantage points. In the future, more web servers can be identified to increase the coverage of Hyperquack.

A motivated actor behind the filter can evade our detection by changing their method of disruption or erasing all identifiable content from their blockpage. While most censors have the capability to evade detection, evasion against our technique is unlikely to be the priority for censors in the short-term because of the following reasons: Blockpages are injected for the purpose of informing users about the reason for blocking and dissuade them from attempting further access. It goes against the purpose of the censor to remove blockpages altogether; Since commercial filters are usually deployed as a black-box, pushing updates to all deployed products to remove blockpages would require significant effort and developer support, and vendors rarely have any incentive to remove trademarks from their blockpages; Changing blockpages is generally easier than removing them altogether. Fortunately, FilterMap would still be able to detect them by identifying new signatures in such cases. Indeed, we believe this to be rare as well, as we observe the same blockpages observed in studies performed several years ago [49], [1], [40].

Moreover, due to accuracy and precision limitations of geolocation databases [28], data from Hyperquack and Quack is labeled with only country level precision. Geolocation information can only identify the vantage point's location rather than the location of the filter deployment, thus our system is not able to determine exactly where on the path the filtering is occurring.

Future work and conclusion By analyzing data from three measurement techniques, Hyperquack, Quack, and OONI, we achieve the most complete view yet on the deployment of censorship filters that respond with blockpages. FilterMap detected filter deployments corresponding to 90 vendors and actors in 103 countries. 20 of these vendors and actors were identified during our longitudinal measurements. All of these attest to the capacity of FilterMap to continuously monitor the evolution of filter proliferation. We find that filters are being used widely to enact censorship, galvanized by high availability and precision in blocking.

A promising future research direction is to use other features of the filter response, such as the certificate returned in HTTPS measurements, to extract signatures and identify filters. FilterMap's analysis techniques can also be used by the censorship measurement community to reduce false positives: measurement artifacts and noise that invariably appear in real-world networks. Moreover, designing effective circumvention tools requires considering both the capabilities of and the methods used by application-level filters. Our system makes it easier for circumvention tool developers to create circumvention strategies based on empirical measurements, by directing them to instances of filter deployment, and providing example domains that trigger them.

The power of filters to implement national-scale Internet censorship has meant that the technology is regulated under export control laws, including the Wassenaar Arrangement, an international mechanism to limit the sale of dual-use technologies [58]. Previous studies revealing the deployments of filtering technologies for censorship have resulted in million-dollar fines for vendor compliance violations, and have helped motivate further regulatory controls [37]. We hope that longitudinal data about filter deployments can help identify those responsible for illegal proliferation of the technology, and provide a basis for more effective enforcement. We intend to maintain FilterMap as a source of longitudinal data for researchers monitoring global censorship, in a format that is readily usable by circumvention developers, advocacy organizations, and regulators seeking to understand and police the proliferation of censorship technologies around the world.

Acknowledgments

The authors are grateful to the shepherd Juan Caballero and to the anonymous reviewers for their constructive feedback. We also thank Bill Marczak, Adam Senft, and Adam Bates for insightful discussions. This material is based upon work supported by the U.S. National Science Foundation under grant CNS-1755841, and by a Google Faculty Research Award.

REFERENCES

[1] G. Aceto, A. Botta, A. Pescapè, N. Feamster, M. F. Awan, T. Ahmad, and S. Qaisar, "Monitoring Internet censorship with UBICA," in *International Workshop on Traffic Monitoring and Analysis*. Springer, 2015, pp. 143–157.

[2] S. Afroz and D. Fifield, "Timeline of Tor censorship," 2007, http://www1.icsi.berkeley.edu/~sadia/tor_timeline.pdf.

[3] Alexa Internet, Inc, "Alexa Top 1,000,000 Sites," 2019, http://s3.amazonaws.com/alexa-static/top-1m.csv.zip.

[4] S. Aryan, H. Aryan, and J. A. Halderman, "Internet censorship in Iran: A first look," in *Free and Open Communications on the Internet (FOCI)*. USENIX, 2013.

[5] H. Asghari, M. Van Eeten, and M. Mueller, "Unraveling the economic and political drivers of deep packet inspection," in *GigaNet 7th Annual Symposium*, 2012.

[6] Beeping Computer, "South Korea is Censoring the Internet by Snooping on SNI Traffic," 2019, https://www.bleepingcomputer.com/news/security/south-korea-is-censoring-the-internet-by-snooping-on-sni-traffic/.

[7] K. Borgolte, C. Kruegel, and G. Vigna, "Meerkat: Detecting website defacements through image-based object recognition," in *24th {USENIX} Security Symposium ({USENIX} Security 15)*, 2015, pp. 595–610.

[8] S. Burnett and N. Feamster, "Encore: Lightweight measurement of web censorship with cross-origin requests," in *ACM SIGCOMM Conference*, 2015, pp. 653–667.

[9] D. Cicalese, D. Z. Joumblatt, D. Rossi, M. O. Buob, J. Augé, and T. Friedman, "Latency-based anycast geolocation: Algorithms, software, and data sets," *IEEE Journal on Selected Areas in Communications*, 2016.

[10] Citizen Lab, "Block test list," 2019. [Online]. Available: https://github.com/citizenlab/testlists

[11] J. R. Crandall, M. Crete-Nishihata, and J. Knockel, "Forgive us our SYNs: Technical and ethical considerations for measuring Internet filtering," in *Ethics in Networked Systems Research*. ACM, 2015. [Online]. Available: https://www.cs.unm.edu/~jeffk/publications/nsethics2015-syns.pdf

[12] J. R. Crandall, M. Crete-Nishihata, J. Knockel, S. McKune, A. Senft, D. Tseng, and G. Wiseman, "Chat program censorship and surveillance in China: Tracking TOM-Skype and Sina UC," *First Monday*, vol. 18, no. 7, 2013.

[13] J. R. Crandall, D. Zinn, M. Byrd, E. T. Barr, and R. East, "Concept-Doppler: A weather tracker for Internet censorship," in *ACM Conference on Computer and Communications Security*, 2007, pp. 352–365.

[14] J. Dalek, L. Gill, B. Marczak, S. McKune, N. Noor, J. Oliver, J. Penney, A. Senft, and R. Deibert, "Planet Netsweeper," 2018, https://citizenlab.ca/2018/04/planet-netsweeper/.

[15] J. Dalek, B. Haselton, H. Noman, A. Senft, M. Crete-Nishihata, P. Gill, and R. J. Deibert, "A method for identifying and confirming the use of URL filtering products for censorship," in *Internet Measurement Conference (IMC)*. ACM, 2013.

[16] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2009, pp. 248–255.

[17] L. Dixon, T. Ristenpart, and T. Shrimpton, "Network traffic obfuscation and automated Internet censorship," *IEEE Security & Privacy*, vol. 14, no. 6, pp. 43–53, Nov.–Dec. 2016.

[18] Z. Durumeric, D. Adrian, A. Mirian, M. Bailey, and J. A. Halderman, "A search engine backed by Internet-wide scanning," in *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2015, pp. 542–553.

[19] Z. Durumeric, E. Wustrow, and J. A. Halderman, "ZMap: Fast Internet-wide Scanning and Its Security Applications," in *22nd USENIX Security Symposium*, 2013, pp. 605–620.

[20] DW, "Internet censorship in Africa threatens democracy, economy," 2018, https://www.dw.com/en/internet-censorship-in-africa-threatens-democracy-economy/a-44956169.

[21] R. Ensafi, J. Knockel, G. Alexander, and J. R. Crandall, "Detecting intentional packet drops on the Internet via TCP/IP side channels," in *International Conference on Passive and Active Network Measurement*. Springer, 2014, pp. 109–118.

[22] R. Ensafi, P. Winter, A. Mueen, and J. R. Crandall, "Analyzing the Great Firewall of China over space and time," *Proceedings on Privacy Enhancing Technologies*, 2015.

[23] Eric Rescorla, Kazuho Oku, Nick Sullivan, and Christopher A. Wood, "Encrypted Server Name Indication for TLS 1.3," March 2019, Work in Progress.

[24] M. Ester, H.-P. Kriegel, J. Sander, X. Xu *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *KDD*, vol. 96, no. 34, 1996, pp. 226–231.

[25] A. Filastò and J. Appelbaum, "OONI: Open Observatory of Network Interference," in *2nd USENIX Workshop on Free and Open Communications on the Internet (FOCI)*, 2012.

[26] Freedom House, "Freedom on the net 2018," 2018, https://freedomhouse.org/report/freedom-net/freedom-net-2018.

[27] K. Fu, C. Chan, and M. Chau, "Assessing censorship on microblogs in China: Discriminatory keyword analysis and the real-name registration policy," *IEEE Internet Computing*, vol. 17, no. 3, pp. 42–50, 2013.

[28] M. Gharaibeh, A. Shah, B. Huffaker, H. Zhang, R. Ensafi, and C. Papadopoulos, "A look at router geolocation in public and commercial databases," in *Proceedings of the 2017 Internet Measurement Conference*, ser. IMC '17. New York, NY, USA: ACM, 2017, pp. 463–469. [Online]. Available: http://doi.acm.org/10.1145/3131365.3131380

[29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[30] A. Houmansadr, C. Brubaker, and V. Shmatikov, "The parrot is dead: Observing unobservable network communications," in *2013 IEEE Symposium on Security and Privacy*. IEEE, 2013.

[31] ICLAB, "ICLAB: Internet censorship lab," https://iclab.org.

[32] "Internet Census 2012 Port scanning /0 using insecure embedded devices," 2012. [Online]. Available: http://census2012.sourceforge.net/paper.html

[33] B. Jones, R. Ensafi, N. Feamster, V. Paxson, and N. Weaver, "Ethical concerns for censorship measurement," in *Ethics in Networked Systems Research*. ACM, 2015. [Online]. Available: https://www.icir.org/vern/papers/censorship-meas.nsethics15.pdf

[34] B. Jones, T.-W. Lee, N. Feamster, and P. Gill, "Automated detection and fingerprinting of censorship block pages," in *Proceedings of the 2014 Conference on Internet Measurement Conference*, ser. IMC '14. ACM, 2014.

[35] J. Knockel, J. R. Crandall, and J. Saia, "Three researchers, five conjectures: An empirical analysis of TOM-Skype censorship and surveillance," in *FOCI*, 2011.

[36] B. Marczak, J. Dalek, S. McKune, A. Senft, J. Scott-Railton, and R. Deibert, "Bad Traffic: Sandvine's PacketLogic Devices Used to Deploy Government Spyware in Turkey and Redirect Egyptian Users to Affiliate Ads?" Citizen Lab, University of Toronto, Tech. Rep., 2018. [Online]. Available: https://citizenlab.ca/2018/03/bad-traffic-sandvines-packetlogic-devices-deploy-government-spyware-turkey-syria/

[37] M. Marquis-Boire, J. Dalek, S. McKune, M. Carrieri, M. Crete-Nishihata, R. Deibert, S. O. Khan, H. Noman, J. Scott-Railton, and G. Wiseman, "Planet Blue Coat," 2013, https://citizenlab.ca/2013/01/planet-blue-coat-mapping-global-censorship-and-surveillance-tools/.

[38] A. McDonald, M. Bernhard, L. Valenta, B. VanderSloot, W. Scott, N. Sullivan, J. A. Halderman, and R. Ensafi, "403 Forbidden: A Global View of CDN Geoblocking," in *ACM Internet Measurement Conference*, 2018.

[39] A. Narayanan and B. Zevenbergen, "No Encore for Encore? Ethical questions for web-based censorship measurement," 2015, https://techscience.org/a/2015121501/.

[40] H. Noman and J. C. York., "West censoring east: The use of western technologies by middle east censors, 2010-2011," 2011, https://opennet.net/sites/opennet.net/files/ONI_WestCensoringEast.pdf.

[41] OpenNet Initiative, "OpenNet Initiative," https://opennet.net/.

[42] OpenNet Initiative, "Jordan," August 2009, https://opennet.net/research/profiles/jordan.

[43] P. Pearce, R. Ensafi, F. Li, N. Feamster, and V. Paxson, "Augur: Internet-wide detection of connectivity disruptions," in *38th IEEE Symposium on Security and Privacy*, May 2017.

[44] P. Pearce, B. Jones, F. Li, R. Ensafi, N. Feamster, N. Weaver, and V. Paxson, "Global measurement of DNS censorship," in *26th USENIX Security Symposium*, Aug. 2017.

[45] PeeringDB, "Peeringdb," 2018, https://www.peeringdb.com/.

[46] W. Scott, T. Anderson, T. Kohno, and A. Krishnamurthy, "Satellite: Joint analysis of CDNs and network-level interference," in *2016 USENIX Annual Technical Conference (USENIX ATC 16)*, 2016, pp. 195–208.

[47] Shodan, "Shodan search engine," 2013, https://shodan.io.

[48] R. Singh, R. Nithyanand, S. Afroz, P. Pearce, M. C. Tschantz, P. Gill, and V. Paxson, "Characterizing the nature and dynamics of tor exit blocking," in *26th USENIX Security Symposium*, 2017. [Online]. Available: https://www.usenix.org/conference/usenixsecurity17/technical-sessions/presentation/singh

[49] The Citizen Lab, "The citizen lab," 2019. [Online]. Available: https://citizenlab.ca/about/

[50] The OONI Tor Project, "OONI Explorer," https://explorer.ooni.torproject.org/world.

[51] The OONI Tor project, "Risks: Things you should know before using ooniprobe ," https://ooni.torproject.org/about/risks.

[52] The Tor Project, "OONI: Open observatory of network interference," https://ooni.torproject.org/.

[53] M. C. Tschantz, S. Afroz, S. Sajid, S. A. Qazi, M. Javed, and V. Paxson, "A bestiary of blocking: The motivations and modes behind website unavailability," in *8th USENIX Workshop on Free and Open Communications on the Internet (FOCI 18)*. Baltimore, MD: USENIX Association, 2018. [Online]. Available: https://www.usenix.org/conference/foci18/presentation/tschantz

[54] UNHRC, "Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression," 2019, https://documents-dds-ny.un.org/doc/UNDOC/GEN/G19/148/76/PDF/G1914876.pdf?OpenElement.

[55] Valentin Weber, "The Worldwide Web of Chinese and Russian Information Controls," September 2019, https://ctga.web.ox.ac.uk/files/theworldwidewebofchineseandrussianinformationcontrolspdf.

[56] B. VanderSloot, A. McDonald, W. Scott, J. A. Halderman, and R. Ensafi, "Quack: Scalable Remote Measurement of Application-Layer Censorship," in *USENIX Security Symposium*, 2018.

[57] Vice, "Netsweeper removes alternate lifestyle category," 2019, https://motherboard.vice.com/en_us/article/3kgznn/netsweeper-says-its-stopped-alternative-lifestyles-censorship.

[58] "The Wassenaar Arrangement On Export Controls for Conventional Arms and Dual-Use Goods and Technologies," https://www.wassenaar.org/, 1996.

[59] N. Weaver, C. Kreibich, M. Dam, and V. Paxson, "Here be web proxies," in *International Conference on Passive and Active Network Measurement*, 2014.

[60] N. Weaver, R. Sommer, and V. Paxson, "Detecting Forged {TCP} Reset Packets," in *Proceedings of the Network and Distributed System Security Symposium, NDSS 2009, San Diego, California, USA*. The Internet Society, 2009.

[61] P. Winter and S. Lindskog, "How the Great Firewall of China is blocking Tor," in *Free and Open Communications on the Internet (FOCI)*. USENIX, 2012.

[62] T. K. Yadav, A. Sinha, D. Gosain, P. K. Sharma, and S. Chakravarty, "Where The Light Gets In: Analyzing Web Censorship Mechanisms in India," in *Proceedings of the Internet Measurement Conference 2018*, ser. IMC '18, 2018.

[63] J. York, "Websense bars Yemen's government from further software updates," 2009, https://opennet.net/blog/2009/08/websensebars-yemens-government-further-softwareupdates.

[64] T. Zhu, D. Phipps, A. Pridgen, J. R. Crandall, and D. S. Wallach, "The velocity of censorship: High-fidelity detection of microblog post deletions," in *USENIX Security Symposium*, 2013, pp. 227–240.

[65] J. Zittrain and B. Edelman, "Internet filtering in China," *IEEE Internet Computing*, vol. 7, no. 2, pp. 70–77, 2003.

APPENDIX

A. Filtering Technology Deployments

Tables VII, VIII, IX, and X show the different kinds of filters detected by FilterMap, the country of deployment, and the datasets that contained the presence of these filters. We use the following symbol notation: An '*' on top of the country name indicates it was discovered using the HTTP dataset, '@' using the HTTPS dataset, '-' using the Echo dataset, '+' using the longitudinal HTTP dataset, '=' using the longitudinal HTTPS dataset, 'o' using the longitudinal Echo dataset, and 'n' using the OONI dataset.

National Firewall	Countries
Bahrain	Bahrain^{*+n}
Iran	Iran^{*-+n}
Saudi Arabia	Saudi Arabia^{*n}
Republic of Korea	Republic of Korea^{-o}

TABLE VII: **Filters with Government blockpages** ◇

Organization with filter deployment	Countries
Brazilian Federal District Government	Brazil^{*+}
Sun TV Network	India^{*+}
AUIS	Iraq^{*+}
Gyeonsang University	Republic of Korea$^{*@}$
Elko	Latvia^{+}
National University Singapore	Singapore^{*+}
Northwestern University	United States^{*+}
Uniminuto	Colombia^{+}
Pustekkom	Indonesia$^{@+=}$
Itgrad	Russia^{-o}

TABLE VIII: **Organizational filter deployments** ◇

Countries	ISP with filter deployment
Ivory Coast	MTN^{*+}
Iceland	STEF^{*+}
Bahrain	VIVA^{*+}
Mauritius	Airtel^{+}
India	Court Order India^{*-+on}
Kyrgyzstan	Elcat*
Saudi Arabia	STCS^{*+}
Yemen	TeleYemen^{*+}
Kuwait	Zain ISP^{n}
Colombia	ERT^{-o}
Pakistan	Wi-tribe^{o}
Republic of Korea	SKT^{n}
United Arab Emirates	Etisalat^{n}
Belgium	Telenet^{n}
Russia	Convex^{-o}, RSVO^{-o}, Piter-Telekom^{-o}, Wiland^{-o}, Kamenkstel^{-o}, Avantel^{-o}, Orion Net^{-o}, Sivash^{-}, Strela Telecom^{-}, Infolink^{-o}, Intertax^{-o}, East Media^{-o}, Sky@Net^{-o}, Sevstar^{-o}, Altegrosky^{-o}, DTEL^{-o}, Goodline^{-on}, Dianet^{-o}, Maglan^{-o}, Skynet^{-o}, Sibitex^{-o}, Novotelecom^{-on}, Yota^{-o}, DSI^{-o}, Kristel^{o}, ITNet^{o}, Westlan^{o}, UGMK-Telecom^{o}, Spacenet^{o}, ACME^{o}, Iterika^{o}, Mosnet^{o}, Metroset^{o}, Redcom^{o}, Bashtel^{o}, tscrimea^{o}, IKS^{o}, Divo^{o}, Beeline^{n}, MTS^{n}, Flex^{n}

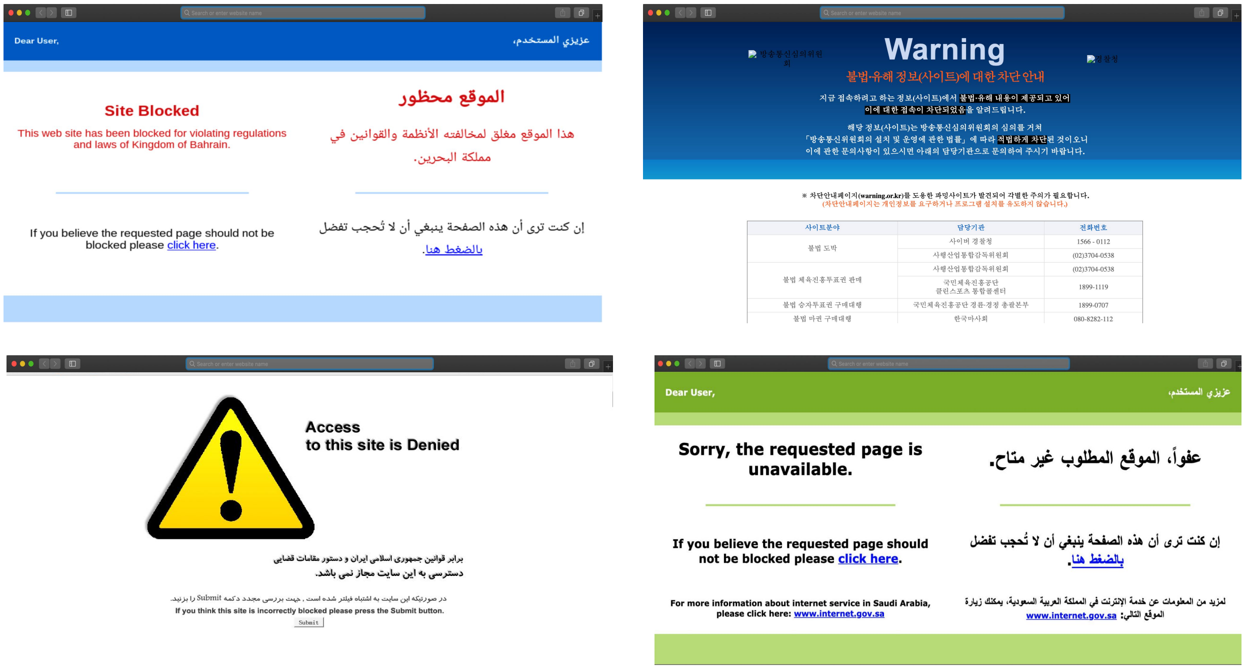
TABLE IX: **Filters with ISP blockpages** ◇

B. Government blockpages

Figure 10 shows the Government blockpages identified by FilterMap.

Filter Vendor	Manufactured in	Countries
Allot	Israel	India^{-}
Barracuda	United States	Estonia*
CacheFlow	United States	Luxembourg$^{*@+=}$
Cisco	United States	Bhutan^{*+}, Pakistan*, Kuwait^{*+}, Mongolia^{*+}, Kyrgyzstan^{*+}, Palestine^{*+}, Qatar^{*+}, United Arab Emirates^{*+}, Iran^{*+}, Kazakhstan^{*+}, Thailand^{-o}, Taiwan^{-o}, India$^{=}$, Austria^{*+}, Croatia^{*+}, Macedonia^{*+}, Belarus^{*+}, Romania^{*+n}, Greece^{*+}, Sweden^{*+o-}, Iceland^{+}, Germany^{n}, Spain$^{=}$, Benin^{*+}, Libya^{+}, Botswana^{+}, United States^{-on}, Canada^{+o}
Fortinet	United States	Saudi Arabia^{*+}, Israel^{*+}, Hashemite Kingdom of Jordan^{*+}, Iraq^{*+}, Turkey$^{*+=}$,Singapore^{*-+o}, Armenia^{*+}, Indonesia^{*+}, Kuwait$^{*@+}$, Oman^{*+}, Taiwan^{*-+o}, Malaysia^{-o}, Palestine$^{*+=}$, Japan^{-o}, Syria^{n}, Republic of Korea^{-o}, Thailand^{-o}, Hong Kong^{-on}, India^{-on}, Philippines^{n}, Lebanon^{n}, Iran^{+}, Republic of Moldova^{*+}, Spain$^{*@+=on}$, Germany^{*+n}, Sweden^{n}, Ireland^{n}, Belarus^{n}, Italy^{n}, Russia^{n}, Hungary*, Luxembourg*, Belgium^{+n}, Ukraine^{n}, Romania^{n}, Republic of Lithuania$^{+=}$, United Kingdom^{n}, France^{on}, Somalia^{*+}, Angola$^{*@+}$, Mali^{*+}, Cameroon^{*+}, South Africa^{*+n}, Egypt^{-}, Gabon^{+}, Zimbabwe^{n}, Zambia^{n}, Kenya^{+n}, Canada^{*+n}, Dominican Republic^{n}, Panama^{n}, Guatemala^{n}, Mexico^{*-+on}, United States^{-on}, Colombia$^{*@n}$, Argentina^{*+} Venezuela^{*+}, Brazil^{-o}, Australia$^{=}$, New Zealand^{n}
IBM QRadar	United States	Saudi Arabia*, India^{*+}
Juniper	United States	Palestine^{*+}, Greece^{*+}
Palo Alto	United States	Malaysia^{*-+on}, United Arab Emirates$^{*@+}$, Thailand^{*-o}, Kuwait^{*-+o}, Mongolia*, Qatar*, Indonesia^{*+o}, Bangladesh^{*+}, Bahrain^{*+}, Saudi Arabia^{*+}, Nepal^{*+}, Hong Kong^{-}, Kazakhstan^{n}, Taiwan^{-on}, Armenia^{+}, Singapore^{n}, Pakistan^{o}, Republic of Korea^{o}, Switzerland^{*+}, Iceland^{*+}, Cyprus*, France^{*+n}, Poland^{*n}, Belarus^{n}, Slovenia^{*+}, Russia^{n}, Croatia*, Latvia*, United Kingdom^{n}, Ireland^{n}, Norway^{n}, Netherlands^{n}, Belgium^{n}, Finland*, Sweden^{*n}, Germany^{*n}, Greece*, Austria*, Italy$^{*@+=n}$, Spain^{-n}, Bulgaria^{o}, Burkina Faso^{+}, Cabo Verde^{*+}, Zambia^{n}, Algeria^{n} , Burundi^{n}, Comoros^{*+}, Puerto Rico*, Canada^{*+n}, Mexico^{*+}, United States^{*-+on}, Brazil^{*+n}, Ecuador^{*-}, Bolivia^{n}, Colombia*
Senhua	China	China*
SmartxFilter	Republic of Korea	Republic of Korea^{+}
SonicWall	United States	Syria^{n}, France^{*+n}, Belarus^{n}, Germany^{n}, United Kingdom^{n}, Russia^{n}, Uganda^{n}, Canada^{n}, United States^{-on}, Brazil^{n}, Bolivia*
Squid	United States	Israel*, Singapore*, Syria^{n}, Turkey^{+}, Belarus^{n}, Germany^{n}, France^{n}, United Kingdom^{n}, Russia^{n}, South Africa^{*+}, Uganda^{n}, United States^{*+n}, Brazil^{n} , Ecuador^{*+}
Sucuri	United States	China^{n}, India^{n}, Iraq^{n}, Malaysia^{n}, Taiwan^{n}, Vietnam^{n}, Belarus^{n}, Austria^{n}, Switzerland^{n}, Germany^{n}, Spain^{n}, France^{n}, Netherlands^{n}, Poland^{n}, Romania^{n}, Russia^{n}, Sweden^{n}, Slovenia^{n}, United Kingdom^{n}, Italy^{n}, Ireland^{n}, Ukraine^{n}, Egypt^{n}, Uganda^{n}, Canada^{n}, Mexico^{n}, United States$^{+=n}$, Brazil^{n}, Argentina^{n}, Ecuador^{n}
VAS Experts	Russia	Russia^{-o}
WatchGuard	United States	Mongolia*, Armenia*,Germany*, Greece$^{*@}$, Finland*, Russia^{n}, Italy*, Belgium*, Serbia*, United Kingdom^{n}, Morocco*, Tunisia^{n}, Costa Rica*, Mexico^{n}, Puerto Rico^{*+}, United States^{n}, Ecuador*, Chile*

TABLE X: **Commercial filtering technologies**: Manufacturer's location based on the company's headquarters. ◇



Dear User,

عزيزي المستخدم،

Site Blocked

This web site has been blocked for violating regulations and laws of Kingdom of Bahrain.

الموقع محظور

هذا الموقع مغلق لمخالفته الأنظمة والقوانين في مملكة البحرين.

If you believe the requested page should not be blocked please click here.

إن كنت ترى أن هذه الصفحة ينبغي أن لا تُحجب تفضل بالضغط هنا.

Warning

불법·유해 정보(사이트)에 대한 차단 안내

지금 접속하려고 하는 정보(사이트)에서 불법·유해 내용이 제공되고 있어
이에 대한 접속이 차단되었음을 알려드립니다.

해당 정보(사이트)는 방송통신심의위원회의 심의를 거쳐
「방송통신위원회의 설치 및 운영에 관한 법률」에 따라 적법하게 차단된 것이오니
이에 관한 문의사항이 있으시면 아래의 담당기관으로 문의하여 주시기 바랍니다.

※ 차단안내페이지(warning.or.kr)를 도용한 파밍사이트가 발견되어 각별한 주의가 필요합니다.
(차단안내페이지는 개인정보를 요구하거나 프로그램 설치를 유도하지 않습니다.)

사이트분야	담당기관	전화번호
불법 도박	사이버 경찰청	1566 - 0112
	사행산업통합감독위원회	(02)3704-0538
불법 체육진흥투표권 판매	사행산업통합감독위원회	(02)3704-0538
	국민체육진흥공단 클린스포츠 통합콜센터	1899-1119
불법 승자투표권 구매대행	국민체육진흥공단 경륜·경정 총괄본부	1899-0707
불법 마권 구매대행	한국마사회	080-8282-112

Access
to this site is Denied

برابر قوانین جمهوری اسلامی ایران و دستور مقامات قضایی
دسترسی به این سایت مجاز نمی باشد.

در صورتیکه این سایت به اشتباه فیلتر شده است , جهت بررسی مجدد دکمه Submit را بزنید.
If you think this site is incorrectly blocked please press the Submit button.

Submit

Dear User,

عزيزي المستخدم،

Sorry, the requested page is unavailable.

عفواً، الموقع المطلوب غير متاح.

If you believe the requested page should not be blocked please click here.

إن كنت ترى أن هذه الصفحة ينبغي أن لا تُحجب تفضل بالضغط هنا.

For more information about internet service in Saudi Arabia, please click here: www.internet.gov.sa

لمزيد من المعلومات عن خدمة الإنترنت في المملكة العربية السعودية، يمكنك زيارة الموقع التالي: www.internet.gov.sa

Fig. 10: **Government blockpages:** (Clockwise) (a) Bahrain (b) South Korea (c) Saudi Arabia and (d) Iran ◇