Decentralized Control: A Case Study of Russia

Reethika Ramesh*, Ram Sundara Raman*, Matthew Bernhard*, Victor Ongkowijaya*,
Leonid Evdokimov†§, Anne Edmundson†, Steven Sprecher*, Muhammad Ikram‡, Roya Ensafi*
*University of Michigan, {reethika, ramaks, matber, victorwj, swsprec, ensafi}@umich.edu
‡Macquarie University, †Independent, §leon@darkk.net.ru

***Abstract*—Until now, censorship research has largely focused on highly centralized networks that rely on government-run technical choke-points, such as the Great Firewall of China. Although it was previously thought to be prohibitively difficult, large-scale censorship in decentralized networks are on the rise. Our in-depth investigation of the mechanisms underlying decentralized information control in Russia shows that such large-scale censorship can be achieved in decentralized networks through inexpensive commodity equipment. This new form of information control presents a host of problems for censorship measurement, including difficulty identifying censored content, requiring measurements from diverse perspectives, and variegated censorship mechanisms that require significant effort to identify in a robust manner.**

By working with activists on the ground in Russia, we obtained five leaked blocklists signed by Roskomnadzor, the Russian government's federal service for mass communications, along with seven years of historical blocklist data. This authoritative list contains domains, IPs, and subnets that ISPs have been required to block since November 1st, 2012. We used the blocklist from April 24 2019, that contains 132,798 domains, 324,695 IPs, and 39 subnets, to collect active measurement data from residential, data center and infrastructural vantage points. Our vantage points span 408 unique ASes that control ≈ 65% of Russian IP address space.

Our findings suggest that data centers block differently from the residential ISPs both in quantity and in method of blocking, resulting in different experiences of the Internet for residential network perspectives and data center perspectives. As expected, residential vantage points experience high levels of censorship. While we observe a range of blocking techniques, such as TCP/IP blocking, DNS manipulation, or keyword based filtering, we find that residential ISPs are more likely to inject blockpages with explicit notices to users when censorship is enforced. Russia's censorship architecture is a blueprint, and perhaps a forewarning of what and how national censorship policies could be implemented in many other countries that have similarly diverse ISP ecosystems to Russia's. Understanding decentralized control will be key to continuing to preserve Internet freedom for years to come.

I. Introduction

Network control has long been a goal of nation-states, and the technology to enable that control is cheaper and easier to use than ever. Countries such as China and Iran have been practicing censorship at centralized network choke points for decades, receiving significant global and academic attention as a result [4], [31], [44], [84]. As more citizens of the world begin to use the Internet and social media, and political tensions begin to run high, countries with less centralized networks have also started to find tools to exert control over the Internet. Recent years have seen many unsophisticated attempts to wrestle with decentralized networks, such as Internet shutdowns which, due to their relative ease of execution, have become the *de facto* censorship method of choice in some countries [14], [36], [83]. While some preliminary studies investigating information control in decentralized networks have examined India [89], Thailand [27], Portugal [60], [61], and other countries, there has yet to be an in-depth multifaceted exploration of the specific tools and mechanisms used by governments for decentralized information control as they evolve over time.

Governments seeking to implement a homogeneous national censorship policy can pursue one of two intuitive options. The first is a centralized control that relies on government-run technical choke points with several layers of complexity, a major government investment that requires an overhaul of the country's network topology. The most notorious example of this, the Great Firewall of China, has cost the country hundreds of millions of dollars [20] over two decades. The second option is to pursue censorship through decentralized control, a task that we have until now deemed to be prohibitively difficult: the case of the Heartbleed vulnerability, where it took 3 months for the gradual installation of patches to reduce vulnerability from nearly 12% of top sites to 3% even after direct disclosure to ISP administrators [18], is an example of the difficulty of coordinating ISPs and their policies. Our study questions the assumption that decentralized network control is too technically difficult and expensive to execute.

To our knowledge, no in-depth study has been performed to assess the feasibility of real-time, effective, and homogeneous information control in a decentralized network. Such a study would require measurements from diverse vantage points, such as ISP backbones to data centers and last-mile residential networks, among others. Furthermore, the research also necessitates knowledge of the country in order to determine what topics, like language, religion or politics, governments are most sensitive to: this makes it challenging to build an exhaustive list of blocked websites. Moreover, even distinguishing between censorship and run-of-the-mill network failures is often difficult, so an insight into the intent of the censor is crucial to establishing which events are censorship events. Finally, determining *who* is actually doing the blocking can be difficult: governments, individual ISPs, and even servers themselves may refuse to serve traffic for a variety of reasons, for instance prioritizing certain customers

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.2020.23098
www.ndss-symposium.org

due to their location [47]. A study examining decentralized information control must account for all of these factors to effectively test the hypothesis of whether decentralized networks can be uniformly censored.

While countries such as India, Thailand, and Portugal are also pursuing decentralized control, the largest and most aggressive country to do so is Russia, which accounts for a sixth of Europe's Internet users [35]. Their censorship regime has grown rapidly over the past decade, with the adoption of policies and laws that facilitate control. We spent a year in continuous discussion with in-country Russian activists who helped us obtain five leaked snapshots of the government's official "blocklist" digitally signed by Roskomnadzor, a primary entity in charge of nationwide Russian Internet censorship. This blocklist contains the list of domains, IPs, and subnets that the Russian authorities have required ISPs to block, and each of its daily iterations since November 1st, 2012. While we have limited historical visibility into how faithfully ISPs applied this blocklist, we can analyze its evolution to understand what the government intended to block through the years.

Our collaboration with activists in Russia also helped us gain access to a diverse set of vantage points in the country, where even renting from reliable Russian virtual private server (VPS) providers requires Russian currency and an in-country phone number and address. From these vantage points, we can perform measurements to provide a clearer picture of Russia's decentralized control—what is blocked, how it is blocked, and how much variation there is from one ISP to another. We performed measurements from within Russia from 20 different vantage points provided to us by volunteer activists, following established ethical practices to reduce risk [19], [77], [91]. We augment the data collected in Russia with two remote measurements tools—Quack and Satellite [9], [58], [78]—expanding our measurements to over a thousand vantage points within Russia and enabling us to validate our local measurements.

From our experiments, we observe that even though not all ISPs block content in similar ways, the volume of websites blocked within residential ISPs is uniformly high. Indicating that coordinated information control in countries with decentralized networks is entirely possible; debunking our initial hypothesis. However, the method by which censorship if effected is largely dependent on their network providers; we observe TCP-layer blocking, application-layer blocking facilitated by deep packet inspection, and DNS manipulation, or a combination of these methods. We also observed that residential ISPs are more likely to inject explicit blockpages, which cite the law and/or Roskomnadzor's registry as they are encouraged to do so by Roskomnadzor's guidelines.

We also observe a difference in quantity and method of blocking between the two network perspectives—residential networks and data center networks. This corroborates the insight that in most countries, residential ISPs are subject to different laws and policies for information control. Therefore, an accurate representative view of censorship is achievable only with measurements from a diverse set of vantage points.

The qualities of Russia's information controls are not restricted to Russia. As Yadav et al. note, India is already attempting to implement a similar censorship regime [89]. The United States [8] and Portugal [60], [61] are both moving away from net neutrality (though not without resistance [53]), and the United Kingdom's legal framework for identifying and restricting content is almost identical to Russia's [75].

The growth of decentralized information control can lead to different ISPs implementing censorship differently, which may contribute to the fragmentation of access to online content for users—even for neighbors who happen to subscribe to different providers. In countries such as China that practice relatively monolithic censorship, circumvention developers can optimize and test tools for use anywhere in the country, and both marketing and word-of-mouth can help users find these effective countermeasures. But in countries such as Russia, decentralized information control adds another layer of complexity: a circumvention tool that works for one user may not work for others. We hope that by highlighting this new trend of moving away from filtering at government-run technical choke points towards legally mandated censorship enforced by private ISPs, we can help inform thinking and future work on other countries pursuing more authoritarian network controls.

II. Background and Related Work

Early censorship research focused on countries with more centralized information controls, such as China and Iran [4], [31]. However, new measurement techniques and in-depth studies of countries such as India and Pakistan [54], [89] have observed a move towards a decentralized approach to information control, through both technical and political means. Technical advancements are making it easier for regimes to restrict their citizens' freedoms even in countries without a history of centralized restrictive controls. Russia is a prime exemplar of this trend, and we fear that Russia will provide a model that other less-centralized countries can adapt. In this section, we delineate centralized and decentralized control, discuss past censorship research, and delve into how Russian censorship embodies an alarming trend, all of which helps guide our understanding of the mechanisms that enable increased decentralized control.

Centralized control: Previous work has shown that censorship within China and Iran follows a very centralized information control scheme [4], [31], [44], [88]. This is made possible by their strict control over the network infrastructure within their respective countries. Countries with centralized control over their network can control information in a highly scalable way, and small perturbations to network reachability can have dramatic effects throughout the country. An example of this is the case in which North Korea's only ISP lost its link with China Unicom, cutting off Internet access in the whole country [59]. Censors like this tend to apply an even mix of censorship methods across the entire networking stack. For instance, China blocks Google's public DNS resolver (8.8.8.8) at the IP layer, Tor relays at the TCP layer [22], poisons many DNS queries [3], [42], and blocks sensitive search terms in HTTP traffic flows [13].

Decentralized control: More recently, several countries around the world have been deploying decentralized information control schemes. These countries do not possess control of their networks in the same way as Iran and China do. Rather, their networks mostly consist of autonomously controlled segments owned by commercial or transit ISPs, whose goals may not align with a government regime attempting to restrict information access. Lack of direct ownership by government authorities lowers their ability to unilaterally roll out technical censorship measures, and instead enact controls via law and policy, compelling the network owners to comply. We see control like this in countries such as India [89], Indonesia [29], and the United Kingdom [2], as well as Russia. In each of these cases, governments pass laws requiring ISPs to block content, and ISPs use a variety of disparate censorship methods to achieve this. For instance, Indonesian ISPs heavily rely on DNS manipulation [29], while Indian ISPs use a combination of DNS manipulation, HTTP filtering, and TCP/IP blocking [89]. These factors cause us to worry that restricting the freedom of citizens is now attainable for many countries, and, even worse, that decentralized information control is more difficult to measure systematically and circumvent. Measuring it requires multiple vantage points within the country and multiple detection techniques to provide coverage of ISP blocking policies. Decentralized control also acts as a barrier to circumvention as it makes it difficult for users to discover locally effective tools.

A. Understanding Censorship Studies

We highlight the common challenges and considerations that drive design decisions in the censorship field, as well as the overview of extant censorship measurement studies and techniques. In this background section, we aim to illustrate how decentralized information control makes it more *difficult* to discover and characterize censorship.

1) Censorship Techniques: On a technical level, network censorship is defined as the deliberate disruption of Internet communication. At the physical layer, a simple form of disruption is to simply "unplug the cable", cutting off all network connectivity. This extreme action has happened on several occasions in a handful of countries. Shutdowns generally are easier to implement for ISPs, but also provoke backlash from customers and impact their business. A recent analysis showed that such disruptions affected 10 countries in sub-Saharan Africa over a combined period of 236 days since 2015, at a cost of at least $235 million [14]. Most studies, including this one, focus on several protocols above the physical layer which are common targets for censorship, we expand on them below and explain common *methods* of interference, protocol and packet features that *trigger* the censor, and the censor *action*.

- *Method: TCP/IP Blocking; Trigger: IP address; Action: Filter request or response*—The censor can disrupt communication to individual services or hosts by blacklisting their IP addresses [1]. This is a particularly common, effective, and cheap way to block access to a server hosting undesired content. It can cause significant collateral damage for innocuous sites that happen to be hosted at the same IP address as a blocked site, e.g. blocking of content delivery networks' (CDN) point of presence [11]. This method has historically been used in countries such as Iran and China to block circumvention proxies such as Tor relays [4], [22].
- *Method: DNS Manipulation; Trigger: Hostname; Action: Filter or modify response*—The censor can observe DNS queries or responses containing a sensitive hostname, decide to either fabricate responses that return DNS error codes such as "host not found", non-routable IP addresses, or the address of a server that likely hosts a blockpage. A blockpage is defined as a notice that explains to the user why the content is unavailable. DNS manipulation enables fine-grained filtering, because simply poisoning the cache of a DNS resolver can be circumvented by using alternate DNS resolvers such as Google's (8.8.8.8).
- *Method: Keyword Based Blocking; Trigger: Keyword, Hostname; Action: Filter or inject*—The censor can inspect and understand the content of the HTTP(S) packets to determine whether it contains censored keywords. The trigger may also be sensitive content in the response or the request other than the hostname. If triggered, it can either drop packets, or inject TCP RSTs or a blockpage. Implementing this form of blocking is challenging, as inspecting traffic at line rate is quite resource-intensive. Naive implementations are trivially defeated; for example, Yadav et al. [89] discovered that merely capitalizing keywords that the censor was looking for entirely circumvented application layer blocking. Some protocols such as HTTPS also defeat naive implementations of application-layer blocking, but more sophisticated blockers may man-in-the-middle each connection and strip the encryption or block based on finding the trigger in the SNI (Server Name Indication) which is transferred in plaintext.

We want to acknowledge that this is a brief overview of the common methods of censorship, and with advancements in traffic filtering technology, sophisticated censors may obtain access to more fine-grained controls to effect censorship.

2) Censorship Measurement Challenges: With the knowledge of how common censorship is implemented, researchers need to tailor measurements to detect most if not all known implementations. There have been numerous other censorship studies that focus on a specific country. Examples of these studies include India [89], Thailand [27], China [12], [31], [33], [88], [93], Iran [4], Pakistan [39], [50], and Syria [10]. While recent work has discussed the political history of Russian's blocking of Telegram [45], our work presents the *first in-depth study of Russia's Internet censorship techniques.*

Effectively measuring censorship requires several components. First and foremost, the *"input list"* of domains or IP addresses being tested can dramatically impact results and effectiveness of any study [56]. Citizen Lab maintains several test lists [40], both general lists of sites that are frequently censored world-wide as well as country-specific lists. Hounsel et al. discusses automatically curating a culture-specific input list by analyzing web pages that are censored in China [33], noting that a lack of an authoritative blocklist can make it difficult to ascertain the intent of the censor and therefore obscure not only why certain sites are censored but also whether measurements of those sites indicate censorship. Further, drawing meaningful conclusions about global censorship and comparing countries is only possible at a category level. But identifying the category of a given website is not a trivial problem. The current state of the art is to use services like

Fortiguard [25] but these services often do not work well for websites other than English.

Censorship measurement studies often suffer from the lack of ground truth which is generally used to validate findings. To compensate for this, studies need to establish strong controls from multiple geographically distributed control vantage points. These vantage points need to be in networks that are not influenced by the censorship regime being studied, and by using multiple vantage points we ensure that the controls are free of effects of transient measurement artifacts and noise. These *"control measurements"* are necessary to establish a baseline for the rest of the study.

In order to comprehensively study the extent of censorship in a particular region, we need a set of *"diverse vantage points"* that shed light on a localized view of the network it operates in. The most direct form of measuring Internet censorship involves using data from users or vantage points (machines under the control of the researcher) inside the country of interest [55]. For example, Winter and Lindskog [84] used one vantage point to study Tor reachability in China and Aryan et al. [4] used one vantage point in their study of Iranian censorship. While one or a few vantage points may be sufficient for measuring centralized censorship regimes, decentralized regimes require a diversity of perspectives.

By making requests to sensitive domains or IP addresses, researchers can directly observe responses from censors and this has been useful for in-depth investigation of censorship techniques in specific countries. These techniques—which we refer to as "direct measurements"—are limited in scale, robustness, and reliability. This is in part due to the difficulty in obtaining vantage points and volunteers and further, due to the potential *"ethical burdens"* of connecting to known-censored content on infrastructure that is likely owned by citizens subject to the jurisdiction of the censor being studied.

In recent years, the popularity of remote censorship measurement tools have grown because of their capability to use more vantage points and perform ethical measurements [21], [57], [58], [70], [78]. These tools do not directly control the vantage points they use for measurement, and thus are not useful for in-depth investigatory testing, but perform well for global censorship measurement. Data collected from remote measurement is also highly complementary to direct measurement since they use different techniques and offer different visibility into the network. Together they are able to offer a more complete view of censorship practices.

Due to observed temporal and spatial variability, recent efforts have focused on developing platforms to continuously collect measurement data on global censorship. One successful platform is Tor project's Open Observatory of Network Interference (OONI) [55], which performs an ongoing set of censorship measurements from the vantage points of volunteer participants [24]. Censored Planet [9], another global censorship observatory, performs continuous remote measurements to identify the prevalence of a variety of censorship techniques in real-time, leveraging the techniques discussed in [57], [58], [70], [78].

3) Censorship Measurement Ethical Considerations: It is important to be aware of the ethical considerations censorship studies take to safeguard participants, regardless of whether they have directly participated (e.g. volunteers) or used as remote vantage points (e.g. organizational servers). Volunteers, especially those in less than democratic regimes, face a risk in accessing sensitive websites. In Section IV we provide comprehensive guidelines that we followed for this study in the hope that it benefits other researchers interested in performing similar work.

B. Russian Information Control

So far we have established common mechanisms by which censorship can occur, and challenges in the way of detecting censorship. In this section, we turn our attention to why Russia's censorship regime is such a compelling example of decentralized control, worthy of study. Russia's censorship regime has seen increased activity in the past decade, but recent events have thrust Russia's information controls into the spotlight. In a famous example, Russia's decision in 2017 to block all Telegram traffic had a massive impact on Internet reachability, as the first attempt to censor Telegram simply blocked millions of IP addresses belonging to the CDNs that Telegram was hosted on [45]. The blocking of these IPs resulted in significant collateral damage, with other services hosted on Google and Amazon becoming unreachable [82].

In order to gain insight into the capability of the Russian government to restrict access to content on the Internet within its borders, we began collaborating with activists within Russia. This collaboration was necessary as Russia has a complex regime of government institutions, each of which control one or a few specific topics that ultimately cause sites to be censored. Our interest stems from the fact that the Russian censorship model can be easily adopted by another country with a similar network structure. In fact, as we discuss in Section VII, other countries such as the United Kingdom already have a censorship regime similar to Russia's (albeit less aggressive). Therefore, we hope that the lessons learned from Russia can help hone future censorship research and meet international regulatory needs to ensure global Internet connectivity.

The rest of this section discusses the specific regulatory and historical characteristics that created Russia's censorship regime. This information helped us shape our research questions, which we present in the following section.

Russian Legal Framework: The primary entity in charge of nationwide Russian Internet censorship is called Roskomnadzor (Federal Service for Supervision of Communications, Information Technology, and Mass Media) [66]. Other government bodies may request that Roskomnadzor block sites, often with content directly related to their scope of duty. The full set of illegal subjects are thoroughly documented by a number of normative acts spanning multiple signed federal laws [64].

Roskomnadzor maintains a singular and centralized Internet blocklist,[1] officially called the Registry of Banned Sites. This registry is an implementation of federal law 139-FZ, passed on July 28, 2012. Currently, Roskomnadzor's registry of banned sites is available to the public, although not in its entirety—only singular queries of an IP address or domain are supported, via

[1] However, there is anecdotal evidence that ISPs sometimes receive slightly different versions and at least one account of Crimea having its own blocklist altogether [76].

a web interface protected with a CAPTCHA [64]. Since its creation, the blocklist has grown in size as new laws were passed to enable the censorship of many subject matters.

Russian Technical Framework: Although Roskomnadzor maintains the central registry of banned sites, they are not behind the technical implementation of censorship in Russia (though they do provide guidelines [67]). Upon the identification of a website with illegal content, Roskomnadzor sends notice to the website's owner and hosting provider. If the illegal content is not removed within three days, the corresponding site is added to Roskomnadzor's registry, and all ISPs across Russia are required to block access to websites in this registry. Therefore, the implementation of censorship falls on Russian ISPs. Complying content owners are able to reinstate access to their websites once violating content has been removed [15]. Notably, the specific method of blocking is not specified, which enables ISPs to implement different censorship mechanisms. ISPs that do not comply with censorship orders sometimes incur fines [72].

While the Russian government itself does not directly censor traffic, it has promulgated some mechanisms for enabling its ISPs to censor traffic. Russia has developed deep packet inspection technology called SORM (System of Operative Search Measures) [62] that it requires ISPs operate in their data centers. The interception boxes themselves are constructed by a variety of commodity manufacturers [62], [79]. While SORM is primarily used for surveillance purposes [73], [74], some ISPs also use it for traffic filtering [79].

Leaked Blocklist: While the blocklist used in Russia is not fully available to the public, we obtained a link to the repository that has regular updates dating back 7 years, as well as official copies of the "current" blocklist signed by Roskomnadzor via our work with activists within Russia. We believe this is the first in-depth study of censorship that has been performed on an authoritative blocklist intended to be used for censorship.

III. Experiment Design

Our experiments to measure Internet censorship in Russia must consider the following factors (1) What to test?–An input list of sensitive content that censors in Russia are likely to block, (2) Where to test?–A set of vantage points from where we can test reachability to websites in the input list, and (3) How to test?–How can we infer details about censorship implementation? In this section we describe how we designed our experiments based on each of these considerations.

A. Acquiring the RUssian BLocklist (RUBL)

We worked extensively with activists within Russia to identify what websites the Russian government has been concerned about. This investigation resulted in our discovery of a leaked blocklist repository [63] with over 26,000 commits dating back from November, 2012, when Russian Internet censorship was still in its infancy. This GitHub repository, Zapret, is well-known within the "Digital Rights guardians" community and is rumored to represent frequent snapshots of the daily blocklists received by ISPs.

We also obtained 5 different digitally signed samples of the blocklist that were distributed by Roskomnadzor, shared with us from multiple sources. We verified that these leaked blocklists are authorized by *CN*=Роскомнадзор and *CN*=Единая информационная система Роскомнадзора *(RSOC01001)* which translates to Roskomnadzor, and Unified Information System of Roskomnadzor. These blocklists are identical to what Russian ISPs would receive. We then compared these blocklists to the Zapret counterpart's contemporaneous commits to corroborate the validity of the repository data and found that the Jaccard similarity between these lists were greater than 0.99. We furnish more details of this validation in Appendix A.

We used the digitally-signed blocklist dated April 24, 2019, which we refer to as *RUBL*, as the input list for all our measurements. A single entry in *RUBL* contains any combination of IP addresses, IP subnets, domains, and domain masks (wildcards). We have no knowledge of how and when DNS resolution was done, or even if resolution was done at all. If the intent was to block domains, we do not know how the accompanying IP addresses were obtained, and vice versa. We break *RUBL* into $RUBL_{ip}$, $RUBL_{dom}$, and $RUBL_{sub}$, containing the unique IPs, domains, and subnets respectively, that pass our controls. Since our measurement tools cannot utilize masks, a domain mask `*.domain.com` is replaced with both `domain.com` and `www.domain.com`. In total, *RUBL* contained 324,695 unique IPs, 132,798 unique domains, and 39 mutually exclusive subnets prior to control measurements which we explain in the following section. While we mainly focus on *RUBL*, we also provide historical analysis of the Zapret repository commits from November 19, 2012, to April 24, 2019 in Section VI.

B. Establishing Sound Control Measurements

Prior to running the measurements from Russia, we need to run control tests to remove IP addresses and domains that are not responsive. To that end, we obtained 13 geographically diverse control vantage points outside of Russia: 4 in North America, 4 in Asia, 4 in Europe, and 1 in Australia. To verify responsive domains, we send a HTTP GET request for every domain from every control vantage point using ZGrab [92], an open-source application layer scanner that operates with ZMap [19]. Our ZGrab tests are customized to follow (a maximum of 10) redirects. We also resolve each of the domains from the control vantage point using ZDNS [90], an open-source command-line utility that provides high-speed DNS lookups. If we get a response for both tests on at least one control vantage point, we include it in the final list. This resulted in a list of 98,098 (73.9% of the original list) domains, which we will refer as $RUBL_{dom}$ for the rest of the paper. We characterize $RUBL_{dom}$ further in section V-B.

We test the responsiveness of the IPs and subnets in $RUBL_{ip}$ and $RUBL_{sub}$ by making TCP connections to port 80 from each control vantage point using ZMap. If we receive a SYN-ACK from the IP to at least one of our control vantage points, we include it in the rest of our measurements. This resulted in 121,025 IP addresses (37.2% of the original list). For $RUBL_{sub}$, we excluded 8 subnets out of the total 39 subnets as they didn't have *any* responsive IP addresses. In total, 567,848 IP addresses (77.2%) were reachable out of 735,232 IP addresses in the expanded subnets. These filtered lists are what we will refer to

VP Type	Num. of VPs	Num. of ASes	Num. of ISPs
VPS in Data Centers	6	6	6
Residential Probes	14	13	13
Quack (Echo Servers)	718	208	166
Satellite (Open DNS Resolvers)	357	229	197
Unique Total	1095	408	335

Table I: **Vantage Point Characteristics** ◇

by $RUBL_{ip}$ and $RUBL_{sub}$, respectively. We characterize them further in section V-A.

C. Conducting Direct Measurement

1) Obtaining Vantage Points: We perform measurements from diverse vantage points, including *VPSes* in data centers and *Probes* in residential networks. An overview of the characteristics of all our vantage points is shown in Table I. To increase our measurement coverage, we also conduct remote measurements discussed later on in Section III-D).

- *VPSes in Data Centers:* With help from activists, we obtained six reliable VPSes confirmed to be hosted in Russian data centers, each in a different ISP. We explored obtaining vantage points from over 35 different providers but many of them observed *no* censorship and some were not conducive to measurement. Renting these machines can only be done with Russian currency and an in-country phone number and address.
- *Residential Probes:* With the insight that different information control policies might apply to residential networks versus data center networks, we also conducted measurements from residential networks. We recruited fourteen participants within Russia to run our probe code (the same that was run at the VPSes, adjusted for lower bandwidth). No information about the participants' network was collected, except for the IP address from which the measurement was performed. To recruit participants, we used the established process of OONI [55] and followed the ethical precautions detailed in Section IV. We attempted to recruit participants from diverse networks, leading us to cover thirteen ISPs (two of our probes were in the same ISP).

In total, our direct measurement platform consists of 20 vantage points. With remote measurements, discussed in Section III-D, we perform measurements from well over 1,000 vantage points. With respect to coverage within Russia, our vantage points are in 408 unique ASes that control ≈65% of Russian IP address space, according to Censys [17].

2) Identifying Censorship Methods: With an established measurement platform and the $RUBL_{dom}$, $RUBL_{ip}$, and $RUBL_{sub}$ lists, we investigate the following: For a given IP address or domain, determine whether it is being blocked; if yes, determine how the blocking is performed. We focus on three common types of blocking: TCP/IP blocking, DNS manipulation, and keyword based blocking based on deep packet inspection. DNS manipulation and keyword based blocking can actuate censorship explicitly by returning a blockpage, or implicitly by forcing a timeout or returning a TCP RST.

Detecting TCP/IP Blocking: We use ZMap to attempt a TCP handshake with each IP address in $RUBL_{ip}$ and in the expanded $RUBL_{sub}$ list. Running this test produces a set of IP addresses that successfully responded to our TCP SYN packet with a TCP SYN-ACK packet. Any IP addresses that do not respond are considered to be blocked, since these IP address were responsive in our control measurement phase.

Detecting Resets and Timeouts: Some censors, when observing an undesirable keyword, drop the packet that forces the connection to timeout or reset the TCP connection. To detect this, we request each domain in $RUBL_{dom}$ interspersed with benign domains such as `example.com` by locally resolving the domain on the vantage point and attempting a HTTP GET request for the domain. This is to ensure that this behavior is not due to transient network errors. If the tests for the benign domains succeed but $RUBL_{dom}$ domains fail, we classify this as censorship due to resets or timeouts, based on the error type received during our test.

Detecting DNS and Keyword Based blocking: More typically when a censored domain is requested, ISPs that employ this method of blocking respond with a blockpage. Detecting blockpages from other unexpected error pages such as server-side blocking errors (e.g. HTTP status code 403), and page not found errors (e.g. status code 404) is not a trivial task. There have been multiple blockpage detection methods proposed in previous work to reduce manual effort [37], [47].

Building on the methodology from Jones et al. [37], our blockpage detection algorithm works as follows: we apply single-link hierarchical agglomerative clustering to HTML web pages to detect blockpages. We extract representative unigrams and bigrams from the clusters under the assumption that pages known from anecdotal sources [7] to contain Russian phrases equivalent to "Access Restricted" and "Roskomnadzor" are usually blockpages, while other sites would not normally contain this kind of language. This is further confirmed by Rozkomnadzor's own recommendations for blockpage content [69].

Using these representative unigrams and bigrams, we manually create regular expressions to match known blockpages. We then validate these regular expressions by grouping pages with the exact same content. We verify that the groups with pages matching the regular expressions contain only blockpages (no false positives). Since ISPs typically return the same blockpage for every censored domain, the *groups* that do not match any regular expressions are not likely to be blockpages, which we manually confirm to eliminate false negatives.

We designed tests that use $RUBL_{dom}$ as the input to characterize DNS and keyword based blocking by employing the decision logic laid out in Figure 1. We explain each test and provide a walk through of the flowchart below.

Test 1: For every domain in $RUBL_{dom}$, we send a GET request from all of our vantage points within Russia, allowing the domain to locally resolve. For all responses that did not contain an error (resets and timeouts categorized and treated separately), we check whether the returned web page matches at least one of the blockpage regular expressions, and if so classify them as "blocked". If this first request is not "blocked", we determine that the domain is not censored. If the request is

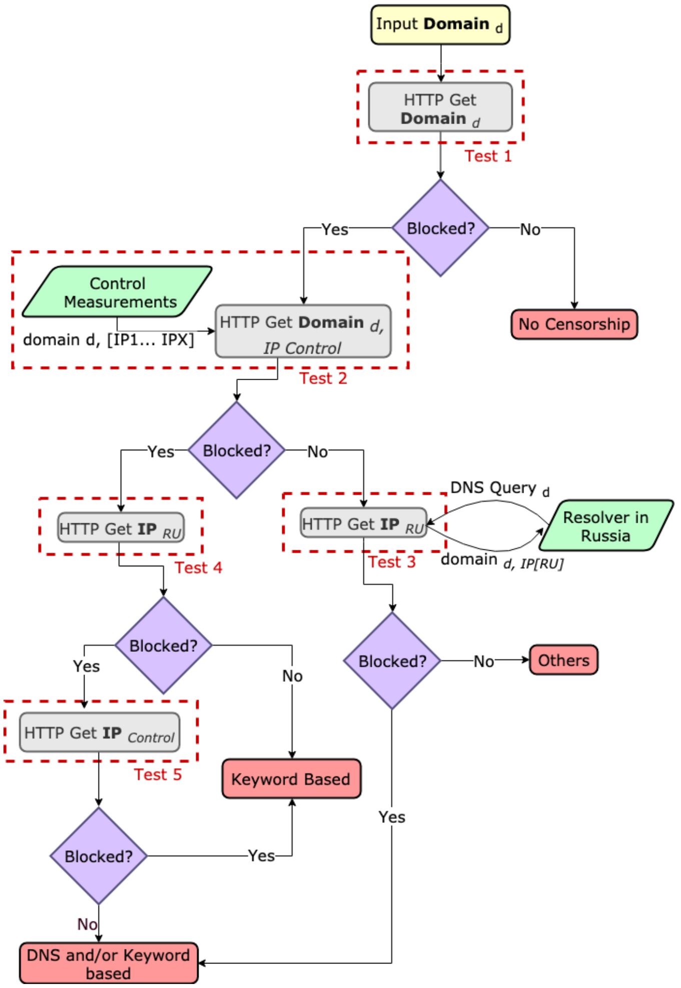


Figure 1: **Decision graph for detecting DNS and Keyword Based blocking**—Four requests are issued: using the domain resolved from a local DNS resolver, using the domain and control IP resolved from every control vantage point, using just the IP resolved in Russia, and using just the IP resolved in controls. We decide whether the request is blocked based on whether the HTML response matches the blockpage regular expressions. ⋄

blocked, we must identify the method of blocking using the results of the following tests.

Test 2: We make another HTTP GET request for the domain, this time using the domain and every unique IP that the domain resolves to in each of the control vantage points. We then pass the web page from the response to our blockpage detection algorithm.

Test 3: If the web page from this Test 2 is not blocked, we look at the result of a GET request for just the IP of the domain resolved in Russia (without the domain name). If the response is classified as blocked from our blockpage detection algorithm, we only know that the domain is either blocked at the application layer by keyword based filtering (if the Russian IP actually points to the site), DNS poisoning (if the Russian IP does not point to the site), or both (if the Russian IP does not point to the site but a blockpage was injected before the connection could reach the poisoned address). If the response is not blocked from this third request, we classify the type of blocking as "Others". Upon investigating what falls under this category, we observed that there are instances where a combination of DNS and TCP/IP blocking is applied, i.e. the actual website is not accessible from the vantage point, even though a blockpage was not received; the reasons may be that the connection was reset or DNS resolution failed every time.

Test 4: If Test 2 is blocked, we look at the result of the GET request with only the IP address resolved from Russia (the same as Test 3), and observe the response. If this is not blocked we can safely conclude that the blocking was only triggered by the presence of the domain name in the request, and thus was blocked at the application layer by keyword based blocking.

Test 5: If Test 4 is blocked, we look at results from the final GET request with only the IP address that was resolved from the control machines. If this request is blocked, we can again definitively declare keyword based blocking, based on some keyword in the *response* from the site also acting as the trigger. If it is not blocked, we can only be certain that it is either DNS manipulation, keyword based blocking, or both.

In cases where we are unable to distinguish keyword based blocking and DNS manipulation we compare the resolved IPs in the Russian vantage points to the resolved IPs in our controls and the answers which are deemed "Not Blocked" in Satellite. The results of this experiment are described in Section VI.

D. Conducting Remote Measurement

Our direct measurements provide a high-fidelity, in-depth view of Russian information control, particularly from the data center and residential network perspectives. However, acquiring these vantage points is quite resource intensive, and our measurements are inherently limited by the number of vantage points we can obtain. To complement this data, and to determine whether our direct measurements are representative, we use two remote measurement tools: Satellite [58], [70] and Quack [78]. Remote measurement tools such as Satellite and Quack use the behavior of existing Internet protocols and infrastructure to detect censorship, i.e. researchers do not need to obtain access to vantage points but just interact with remote systems to learn information about the network. Satellite remotely measures DNS manipulation using open DNS resolvers and Quack detects application-layer blocking triggered on HTTP and TLS headers using Echo servers. These remote measurements select only vantage points that are part of organizational or ISP infrastructure, hence providing a complementary perspective to direct measurements.

1) Obtaining Remote Vantage Points: With operational help from the Censored Planet team [9], we used 357 open DNS resolvers in Russia located in 229 different ASes (197 unique ISPs), and 718 Echo servers located in 208 different ASes (166 unique ISPs). As shown in Table I, this increases our coverage considerably. We annotate the vantage point locations with the Maxmind GeoIP2 database [46], and find the AS information through RouteViews data [68].

2) Identifying Censorship: On our behalf, the Censored Planet team performed Satellite and Quack using $RUBL_{dom}$ based on the techniques described in [78] and [70]. Both tools have their own methods to label a domain as being "manipulated" or "blocked". Satellite creates an array of five metrics to compare the resolved IP against: Matching IP, Matching HTTP content hash, Matching TLS certificate, ASN, and AS Name. If a response fails all of the control metrics, it is classified as blocked. Quack first makes an HTTP-look-alike request to port 7 of the Echo server with a benign domain (`example.com`). If the vantage point correctly echoes the request back, Quack then requests a sensitive domain. Quack makes up to four retries of this request in case none of the requests are successfully echoed back. If the vantage point fails for all 4 requests, Quack tries requesting a benign domain again to check whether the server is still responding correctly. If so, the failure to echo back the sensitive domain is attributed to censorship.

IV. Ethical Considerations

Censorship measurement studies involving active network measurement raise important ethical considerations. Most censorship measurement studies, including ours, aim to trigger censors from various vantage points which might cause risk of retribution from local authorities. Aiming to set a high ethical standard, we carefully designed our experiments to follow or exceed the best practices described in the Belmont [51] and Menlo [16] reports. Before initiating any of the measurements, we consulted with our university's IRB, who determined that we were exempt from regulation but advised us to discuss with the university's General Counsel, which we did. We vetted the risks of our study and shaped our data collection methods through a year of continuous communication with prominent activists within Russia, with colleagues experienced in censorship and measurement research, and with our university's General Counsel.

Gaining background understanding of the laws of the country is imperative to designing ethical measurements. Prior to engaging with us, our activist collaborators had been actively participating in open-source projects such as OONI and Tor, and had traveled outside of Russia to present details about Russian censorship in international forums. Their guidance was essential for us to ensure we were aware of Russian law and policy regarding accessing censored content. These collaborators facilitated renting VPSes and running measurement from the residential probes.

Our direct measurements involve sending requests for potentially censored content from vantage points inside Russia. This creates a potential risk to participants who own and control these vantage points. We consulted with our activist collaborators, who assured us that even if the anonymized vantage points, data centers, or ISPs are discovered, there has never been any punitive action on the part of the Russian government or others against entities who do not comply with the blocklist. We then begin the process of obtaining informed consent from participants by customizing the OONI consent form which was drafted by the Harvard Cyberlaw Clinic and attached in the Appendix E). This form documents in detail the measurements performed and data collected and seeks explicit approval. Before our activist collaborators asked participants to run measurements from residential probes, they used our consent form and drafted an email in Russian to solicit explicit consent from the volunteers, who were recruited from a tech-savvy population already involved with activist groups that advocate for Internet freedom.

We obtained our VPSes from commercial VPS platforms, whose operators understand the risk in offering network and computing services. In collecting the data from our VPS platform, we did not subject anyone in Russia (or elsewhere) to any more risk than they would already incur in the course of operating a VPS service.

Our remote measurements seek only vantage points that are not owned or operated by end users and are part of organizational or ISP infrastructure. As in the case of our VPSes and residential probes, there is a possibility that we place the operators of these remote vantage points at risk. Again, there is no documented case of such an operator being implicated in a crime due to any remote Internet measurement research, but we nonetheless follow best practices to reduce this hypothetical risk. From the list of all available open DNS resolvers in Russia, we identify those that appear to be authoritative nameservers for any domain by performing a reverse DNS PTR lookup and only select those resolvers whose PTR begins with the regular expression "ns[0-9]+|nameserver[0-9]". Similarly, we ran Nmap on all the Echo servers in Russia and exclude those whose labels do not indicate an infrastructural machine. Using only infrastructural vantage points decreases the possibility that authorities might interpret our measurements as an attempt by an end-user to access blocked content. Moreover, we initiate the TCP connection and send the sensitive requests, and there is no communication with the actual server where the sensitive domain is hosted. We also set up reverse DNS records, WHOIS records, and a web page served from port 80 on each machine in the networking infrastructure we use to run measurements, all indicating that our hosts were part of an Internet measurement research project.

We also follow the principle of good Internet citizenship and reduce burden on the vantage points by rate limiting our measurements, closing TCP connections, and maintaining only one concurrent connection. Our ZMap and ZGrab scans were conducted following the ethical guidelines proposed by Durumeric et al. [17], [19].

V. Data Characterization

The most recent sample of *RUBL* contains 132,798 unique domains and 324,695 unique IP addresses. It also contains a list of 39 subnets ranging from /24s to /16s. This section characterizes both the full *RUBL* blocklist and the final filtered list obtained after running control measurements described in Section III-B.

A. IPs and Subnets

As mentioned in Section III, we examined the responsiveness of the IPs on the blocklist. Only 121,025 IPs on the blocklist (37.3%) were reachable from our controls. Our control measurements were highly concordant; over 99% of IPs that were reachable at some control vantage point were reachable at all control vantage points. The low rate of responsiveness (37.3%) might be the artifact of our measurement, as these IPs might be alive but not responding on port 80, such as proxies configured on custom ports.

#	Country	IPs	#	Country	IPs
1.	United States	203,107	6.	Russia	6,328
2.	Germany	31,828	7.	Finland	6,057
3.	United Kingdom	25,931	8.	Japan	2,490
4.	Netherlands	16,161	9.	Estonia	2,327
5.	France	8,117	10.	Iran	2,070
Other					19,622
Total					324,038

Table II: **Top ten countries hosting IPs on the blocklist.** ◇

TLD	Domains	CDN	Domains
1. .com	39,274	1. Cloudflare	44,615
2. .ru	11,962	2. App Engine	89
3. .info	5,276	3. Cloudfront	80
4. .net	4,934	4. Incapsula	48
5. .xyz	3,856	5. Akamai	12
	—	In two of the above	47
Others	32,796	No CDN	53,301
Total	98,098	Total	98,098

Table III: **Top five TLDs and CDNs for domains in the blocklist**—.com and .ru are the most popular TLDs. ◇

For the 324,695 unique IPs in the list, we examined their geolocation using the MaxMind Geolite2 [46] database. 324,038 (99.8%) IPs were found in the database. We saw that over 200k IPs (>61%) were located in the US. Somewhat surprisingly, Russia was only the sixth most popular country in which IPs were located as shown in Table II. These IPs spanned over 2,112 Autonomous Systems (ASes) based on RouteViews lookup.

The blocklist also contains 39 subnets, ranging from /16s to /24s. 31 out of 39 of these subnets contain at least one IP reachable to one of our controls. The remaining eight unreachable subnets geolocated to Moscow.

B. Domains

For the 132,798 domains in the list, over 49,583 (37.3%) are `.com` domains and 15,259 (11.5%) are `.ru` domains. As discussed in Section III, 34,404 (25.9%) domains on the blocklist are not responsive, so for the analysis that follows we only focus on the 98,098 responsive domains. `.com` and `.ru` still dominate responsive domains as shown in Table III.

Inspired by McDonald et al. [47], we looked at what CDNs the sites in the blocklist were hosted in, if any. We were able to identify the CDN for 44,797 (45.7%) domains following their methodology. As shown in Table III, an overwhelming majority of domains which were served by a CDN (99.6%) were hosted on Cloudflare, which provides some of its services for free with little vetting of the sites. 47 domains had signs that they used more than one CDN service. In these cases, we counted them as customers of both.

We initially experimented with using the Fortiguard document classification service [25] to categorize domains and ascertain what types of websites are in the blocklist. Unfortunately, the Fortiguard classification was not effective for Russian language domains. Also, a large number of domains—27,858 (28.4%)—were classified into the "Business" category, which did not reveal much information about the services hosted on those domains. Therefore, we developed our *topic modeling* algorithm designed after the technique introduced in Weinberg et al. [81]. Our topic modeling algorithm processes the text received from control measurements, and uses Latent Dirichlet Allocation (LDA) clustering [6] to identify pages with the same topic. To accomplish this, we adopt the following steps:

- *Text Extraction*—From the control measurements, we obtained the HTML responses for all the 98,098 domains. We first filter out all the responses that returned an empty HTML body, have an error code in the status line, or have encoding issues in the server response. This reduced the number of classifiable domains to 70,390 (71.8% of the original list). We then use Python's `Beautiful Soup` library [5] to extract useful text and remove boilerplate text.
- *Language Identification*—The LDA algorithm requires input documents to be in the same language; as described in [81], it detects semantic relationships between words based on the probability of them occurring together within a document. We used Python's `langdetect` library [41] to identify the primary language for each document. Out of 70,390 classifiable documents, 44,270 (62.9%) primarily contained Russian or related Cyrillic text, and 19,530 (27.7%) contained primarily English text. We choose to focus on this portion of the classifiable pages as the other 9.4% contained documents in 42 different languages. We thus reduce our manual effort in labeling topics by only using LDA only twice, once for Russian pages and once for English pages.
- *Stemming*—Before applying the LDA algorithm, we reduce all words to stems using Snowball [71]. We then apply term frequency-inverse document frequency (*tf-idf*) [65] to select terms that occur frequently. We preserved terms whose combined *tf-idf* constitutes at least 90% of the total document.
- *LDA analysis*—We then use LDA for Russian and English documents separately. We used Python's `gensim` [28] and `nltk` [52] libraries for our implementation, and we used all documents for training. We found N=20 topics to be optimal, and α is determined optimally by the library based on the training data.

Using LDA, we obtain 20 topic word vectors from the English documents and 20 topic word vectors from the Russian documents. Two researchers independently labeled the topics by reviewing the top words in each topic. Disagreements were resolved through discussion between the researchers. Many topics were given the same label; as discussed in [81], this is one of known limitations of LDA analysis. We manually merge these topics into 9 categories. Additionally, we manually selected a random subset of documents within each topic cluster and ensured that all the documents belonged to the category they were assigned.

The number of English and Russian documents classified into each category is shown in Table IV. The majority of domains (67.6%) fall into the "Gambling" category, indicating the stringent crackdown of Russian authorities against gambling websites. Our analysis suggests the high number of gambling websites to be an effect of websites quickly cloning to an alternate mirror domain when added to the blocklist. This can be seen by many of the gambling website domains on the blocklist having slight vari-

Category	Num. Russian	Num. English	Total
Gambling	33,097	10,144	43,241
Pornography	5,576	2,821	8,397
Error Page	134	3,923	4,057
News and Political	1,883	No clusters	1,883
Drug Sale	1,811	No clusters	1,811
Circumvention	1,769	No clusters	1,769
Multimedia	No clusters	1,610	1,610
Parking Page	No clusters	601	601
Configuration Page	No clusters	431	431
Categorized Total	44,270	19,530	63,980
Other Language Pages	—	—	10,464
No HTML or Error	—	—	23,654
Total			98,098

Table IV: **Categories of responsive domains obtained using topic modeling**—The second column shows the number of documents in primarily Russian or related Cyrillic languages classified into each category, and the third column shows the same for primarily English language documents. Gambling and pornography websites dominate the blocklist. ◇

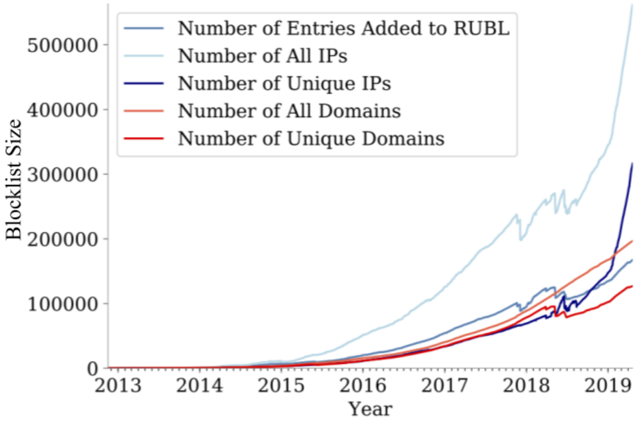


Figure 2: **Evolution of the blocklist over 7 years**—The blocklist has grown rapidly for much of its existence, across all categories of contents. ◇

ations in their names, for example `02012019azino777.ru`, `01122018azino777.ru`, `01042019azino777.ru`, and so on. This also suggests that the blocklist is not actively maintained. Unsurprisingly, pornography websites also feature prominently in the blocklist.

$RUBL_{dom}$ contains news, political, and circumvention websites that feature exclusively Russian-language media (`chechenews.com`, `graniru.org`) and activist websites such as `antikor.com.ua`, which is a self-proclaimed national anti-corruption portal. Some of the pages were also categorized into error pages, parking pages and configuration pages, indicating that these domain owners have moved since being added to the blocklist. These pages are primarily in English because they use templates from popular web server error pages (e.g. Apache, Nginx etc.)

There are a few *caveats* to our topic modelling algorithm. First, the documents we determine as Russian and English may contain text in other languages, but we only choose those documents that are predominantly in either Russian or English. Nevertheless, a significant amount of other language text may lead to miscategorization of some websites. Second, our labeling is primarily based on the top words in each word vector. This may also lead to some pages being categorized incorrectly, but our manual verification did not find any false positives.

VI. Results

We divide this section into four parts: first, we begin with an analysis of the Zapret repository and present data about how it has evolved over time. Then we present results from $RUBL_{dom}$, $RUBL_{ip}$, and finally, $RUBL_{sub}$ measurements.

A. Historical Analysis of Russian Blocklist

We analyze the Russian blocklist's evolution over a seven-year time period, from November 19, 2012, to April 24, 2019 at a daily granularity. Since it may be updated multiple times a day, we utilize only the latest version, which is most often published close to midnight. Any activity of smaller granularity, such as the occasion of an addition or removal of an IP address in a time span of less than 24 hours, is not considered. IP subnets are not included in this analysis, which amount to an approximately additional 26,000 addresses beginning in the middle of 2017 and 16 million addresses beginning in April 2018 due to the banning of Telegram. These addresses are omitted because their inclusion obscures graph clarity due to their significantly greater scale.

As shown by Figure 2, *the size of the blocklist appears to have grown rapidly since its conception in 2012.* The plot shows three size metrics: number of entries, raw number of both IPs and domains, and number of unique IPs and domains. Each of these metrics is cumulative and the drops in the number are due to "removal" of entries, IPs, or domains. Since an entry may contain multiple IPs and domains, the number of IPs and domains far exceeds the number of entries.

An unexpected finding is how the raw number of IPs significantly exceeds the number of unique IPs. This discrepancy can be attributed to potentially unintentional duplication—one IP added to the blocklist because it hosts one domain name may later be entered again for a different domain. Multiple domains may share IPs because of the prevalence of sites hosted on CDNs in the blocklist (as discussed in Section V-B). More details on this analysis can be found in Appendix B.

One important observation is the sharp increase in the number of raw IPs, *unique* IPs, and a moderate increase in the number of unique domains in the past year. This suggests that there is a deliberate effort to increase the accuracy of the list. This is further punctuated by a number of drops in all the metrics in the past year, which suggests that there has been conscious effort put into making the list more meaningful and to avoid repetitions.

B. Characterizing Censorship of $RUBL_{dom}$

As described in Section III, we have six VPSes in data centers and 14 residential probes. Figure 3 shows the type of censorship observed at each vantage point. We divide the rest of this section by vantage point type, in order to highlight the complementary nature of the results from each of them.

1) VPSes in data centers: We observed some amount of censorship at all of our VPSes in data centers. The number

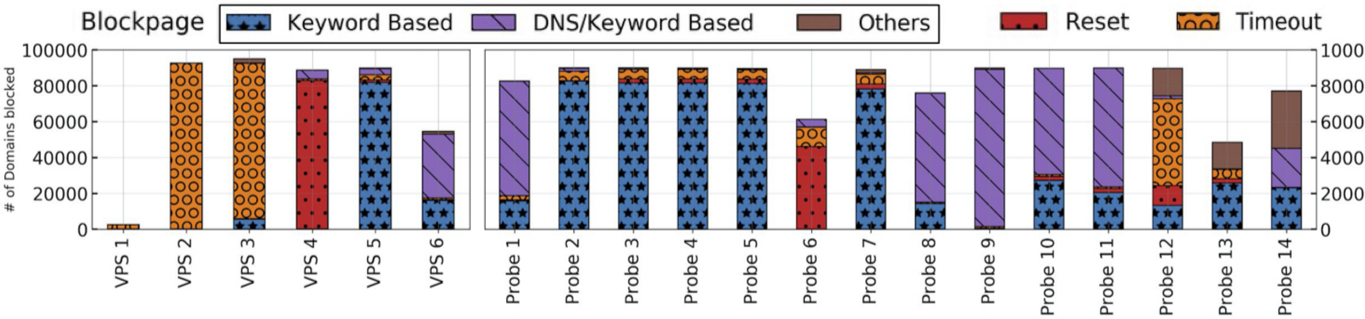


Figure 3: **Testing** ***RUBL$_{dom}$*** **from all vantage points**—The kind of blocking varies between vantage points. VPSes in Data Centers see varying levels of blocking. Residential Probes experience a larger amount of domains blocking, and they also typically receive explicit blockpages. ◇

of domains blocked per vantage point is shown in Figure 3. Four out of six VPSes show that more than 90% of *RUBL$_{dom}$* is blocked, with the highest blocking 96.8% of all *RUBL$_{dom}$* domains.

The censorship method varies between each VPS, confirming our hypothesis that *the lack of prescription of censorship mechanism enables data center network providers to employ any method of censorship.* While most VPSes observe multiple kinds of blocking, one method of blocking typically dominates at each vantage point. For example, VPS 5 and VPS 6 mostly observed blockpages, while VPS 2 and VPS 3 observed more connection timeouts. In VPS 4, we observed that TCP connections were reset when domains in *RUBL$_{dom}$* were requested. We suspect that VPSes observe more than one type of blocking due to content being blocked at different locations along the path to the server, such as at transit ISPs. Content restriction at transit ISPs would cause most content to be blocked across the country, even if ISPs closer to the user do not censor all content in *RUBL*.

2) Residential Probes: Figure 3 shows that residential probes show higher amounts of blocking overall, suggesting that ISPs closer to the user block almost all the domains more uniformly. Nine out of 14 residential probes observe more than 90% of the domains blocked and all of the probes observe at least 49% of the domains blocked.

While VPSes saw high occurrences of timeouts and resets, most residential probes observed a blockpage. We believe this is in part due to the fact that residential ISPs are encouraged by Roskomnadzor's guidelines [67] to cite the law and/or Roskomnadzor's registry and provide explicit information regarding blocking to users. As for the other methods of blocking, we found that Probe 6 predominantly observed a large amount of connection resets and Probe 12 observed a large number of timeouts.

As mentioned earlier, a blockpage is shown to the user when the blocking method is either "Keyword Based" or "DNS/Keyword Based". In the latter the trigger is the hostname but the method of blocking is not clear. In an effort to distinguish between the two methods of blocking, we compare the IPs from domain resolution in the residential probes with the IPs received in domain resolution from all control vantage points and with the answers that were determined as "Not manipulated" in Satellite. The percentage of IPs from each vantage point that does not match any control IP or any resolved IP in Satellite is shown in Figure 4. VPS 6, Probe 9 and Probe 14 observe a large percentage of resolved IPs that do not match any of the control responses. This lends credence to the hypothesis that these three vantage points may be subject to DNS manipulation rather than keyword based blocking. To corroborate this, we investigate all instances of "DNS/Keyword Based" blocking and found that *each of the three vantage points observed a single poisoned IP respectively*. We looked at the content hosted at these three IPs and found a blockpage being returned which can be seen in Figure 10 in the Appendix.

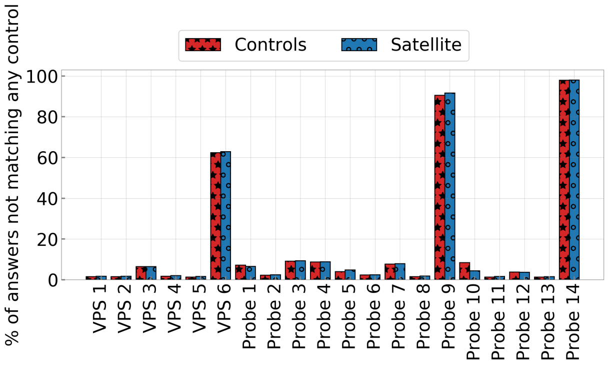


Figure 4: **Answers from DNS resolutions that do not match answers from any control DNS resolutions or Satellite resolutions**—Three vantage points VPS-6, Probe-9 and Probe-14) show signs of DNS manipulation. ◇

We observe blockpages in that was categorized as "Other" specifically in Probes 12, 13, and 14, meaning we could not exactly determine the *method* of blocking. Upon investigation, we saw that Probe 14 received a blockpage when queried with the domain but was unable to retrieve the page when queried with the IP received from control. Considering Probe 14 also sees high IP blocking as shown in Figure 7, we believe Probe 14 observes a combination of DNS and IP blocking. Similarly for Probes 12 and 13, we observe behavior consistent with Keyword Based blocking but the blockpage was unable to load in some cases.

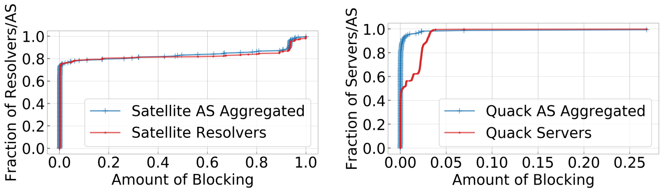


Figure 5: **Fraction of domains blocked at the individual vantage point as well as AS (aggregated) level**—There are some vantage points and ASes that only block little content, while others block comparatively many more domains. The similarity between the lines shows that blocking is happening at the AS level. Our measurements using Satellite observed much more interference compared to Quack measurements. ◇

3) Remote Measurements: We conduct remote measurements for $RUBL_{dom}$ using 357 vantage points for Satellite and 718 for Quack. The CDFs in Figure 5 show the blocking behavior for resolvers in Satellite and echo servers in Quack. There are large variations in the fraction of blocking between vantage points in both Satellite and Quack. There are some vantage points that do not observe any blocking, while others observe a large amount of blocking. Between Quack and Satellite, Satellite observed considerably more blocking, which is in line with at least three of our vantage points that observed large amounts of DNS manipulation. We suspect that many Russian ISPs may not be blocking content on port 7, and hence are not captured by Quack. This is a known shortcoming of Quack by not triggering censors that only act on port 80 and 443. This suggests that one method of circumventing censorship might be serving content over non-standard ports.

Figure 5 also shows the fraction of blocking aggregated at the AS level. The similarity between the two CDFs shows that *blocking does indeed happen at the AS or ISP level.* In Satellite, we observe that more than 70% of vantage points observe little to no blocking, while in Quack 50% of vantage points observe no blocking, and close to 90% observe minor blocking.

In our Quack measurements, we were able to look at the kind of blocking observed at each of the echo servers. Similar to our observations in the VPSes in data centers, some vantage points observe blockpages, many others observe resets and timeouts (more frequently resets), showing that censorship mechanisms vary widely in networks all over Russia.

We looked at the similarity between domains being blocked in our remote vantage points. The pairwise similarity is shown in Figure 6. We see that our observations from the VPSes and residential probe measurements are consistent with remote measurements as well. Both Satellite and Quack see instances of high similarity, which is either because the vantage points see a high percentage of domains blocked (top left) or because vantage points are inside the same ISP (small square stripes along the diagonal line). The large blue portions on both plots show that vantage points which observe little blocking do not see the same domains being blocked.

C. Characterizing Censorship of $RUBL_{ip}$ Measurements

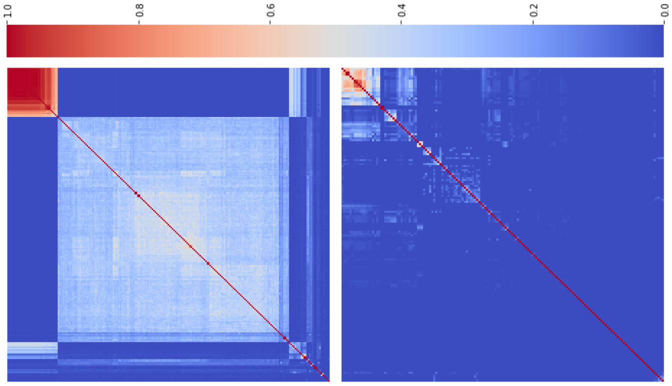


Figure 6: **Pairwise Jaccard similarity of domains blocked in remote measurements**—As in the direct measurements, we observe some similarity between domains blocked in remote measurements (Satellite on the left, Quack on the right) either due to high blocking or vantage points in the same ISP. ◇

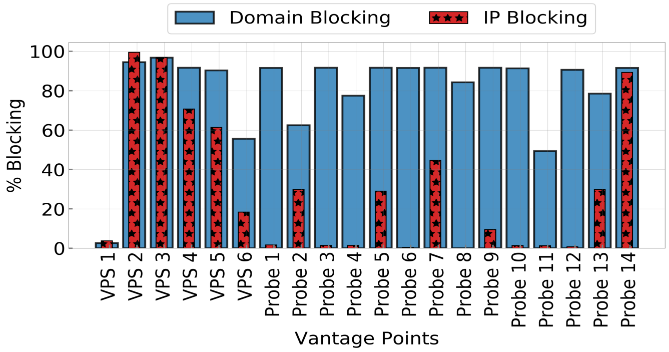


Figure 7: **Blocking by method when testing $RUBL_{ip}$ on VPSes and residential probes**—Data center vantage points observe much higher IP blocking compared to residential probes, where domain blocking is more popular. ◇

We study the extent of blocking of IPs in $RUBL_{ip}$ by analyzing the output of our TCP/IP measurements from both our VPSes and probes. The amount of IP blocking is shown by the red bars in Figure 7. For comparison, we overlay the total percentage of domains blocked in these vantage points as well. Overall, we see a smaller percentage of IPs being blocked compared to domains, which could indicate a desire by the censors to minimize collateral damage (other services hosted on the same IPs would be blocked as well). Alternatively, it could be that residential ISPs do not observe much traffic to IPs, and opt to censor only the traffic they see.

Similar to our observations in $RUBL_{dom}$, we find that there are some vantage points which observe blocking of many IPs, while other vantage points only observe a few blocked IPs. VPSes observe a considerable amount of IP blocking, while the blocking is more sparse in probes. Our experience suggests that *data center VPS providers could also be injecting resets and forcing timeouts to these measurements as well.* In the residential probes, only Probe 14 observes more than 50% of IPs being blocked, while four out of six VPSes observe more than 50% IP blocking. This seems to corroborate the hypothesis that *residential ISPs tend to block the kind of traffic they see more frequently, which is predominantly traffic involving domains.*

Vantage Point	Num. of subnets	Vantage Point	Num. of subnets
VPS 1	2	Probe 1	5
VPS 2	31	Probe 2	27
VPS 3	4	Probe 5	6
VPS 5	5	Probe 9	2
VPS 6	1	Probe 10	5
		Probe 11	5
		Probe 13	2
		Probe 14	6

Table V: **Number of subnets completely blocked by vantage points**—VPS 2 and Probe 2 block almost all of the subnets in $RUBL_{sub}$ completely, while others moderately block subnets. ◇

D. Characterizing Censorship of $RUBL_{sub}$ Measurements

Table V shows the number of subnets that were completely unreachable from our vantage points, omitting the vantage points where at least one IP from each subnet was reachable. Keeping in line with our previous observations, we see that there are some vantage points that block nearly all of the subnets (e.g. VPS 2 and Probe 2) some that block a moderate amount (e.g. VPS 6), and some that do very little blocking (e.g. Probe 12) corroborating our findings in Section VI-C that *different ISPs may prioritize blocking different items in RUBL.*

Similar to our observation in $RUBL_{ip}$, VPSes in data centers observe much higher blocking in $RUBL_{sub}$ compared to residential probes, where only Probe 2 observes a large amount of $RUBL_{sub}$ blocking. Our $RUBL_{ip}$ and $RUBL_{sub}$ study suggest that *most residential ISPs prefer to block using the domain in the request, as opposed to the IP to which users are ultimately connecting to.* Further $RUBL_{sub}$ analysis can be found in Appendix D

VII. Discussion and Conclusion

Russia's move towards more restrictive Internet policies is illuminating in the broader context of tightening information controls around the world. Censorship studies have until this point mostly focused on centralized networks like those in China and Iran; Russia's network however, like that of most countries throughout the world, was shaped gradually by many competitive market forces. The development of effective censors on a decentralized network such as Russia's raises important questions on the future of censorship including in western countries that have not historically favored censorship.

Our study has shown that the implementation of decentralized control breaks the mold of the traditional definition of "censorship": a synchronized and homogeneous process of blocking sensitive content throughout the country. While Roskomnadzor has coordinated blocking across various ISPs that all have independent motives, they are yet to achieve homogeneity in the method of blocking. But with the advent of SORM and the commoditization of censorship and surveillance technology, it is becoming cheaper and easier for ISPs to comply with government demands.

The variegated nature of Russia's censorship regime has significant implications for censorship research moving forward. It is no longer sufficient to perform measurements from one or a few vantage points within the censoring country. Even two end-users in the same physical location may have dramatically different experiences with censors based on their ISP; they would both see very different results than data centers. Measuring the actual impact of censorship also proves difficult: it requires diverse vantage points, including residential network ISPs, infrastructural machines as well as data centers.

Decentralized control also has significant implications for censorship resistance and creates more barriers discovering effective circumvention. As Russia moves to block access to VPNs [48], users will need to rely on more exotic means of circumvention. Since the method of blocking varies between networks, there is increased difficulty in finding locally effective circumvention tools. Techniques like refraction networking [26], [32], [38], [86], [87] or domain fronting [23] may become necessary. In any event, Russia has sparked an arms race in censorship and circumvention, and its effects are likely to be felt around the world.

We have already started to see other large nations begin applying schemes similar to Russia's. In the United States, ISPs have been rolling out DPI boxes over the past decade which can dynamically throttle connections to specific websites, [30], [43], [49] or favor certain content over others [8]. The United Kingdom's censorship model is similar to Russia's, with the government providing ISPs a list of websites to censor [85] and having governing bodies that correspond to various types of censored material [75]. For both the U.S. and the U.K., what this means is not that the current regimes are restricting the volume of information that Russia is, but that the option to follow the same path is cheap, readily accessible.

The same can be said for nations around the world. Portugal has recently been cited for not supporting net neutrality [60], Indonesia recently implemented broader content filtering [34], and India has been ramping up censorship [89]. A recent report [80] finds that Russian information controls and the technology used for surveillance and censorship capabilities are being exported to at least 28 countries. As more countries move towards stricter Internet access, Russia's model for censorship may become more commonplace, even in countries with a tradition of freedom of expression on the Internet.

In conclusion, Russia's decentralized information control regime raises the stakes for censorship measurement and resistance. Its censorship architecture is a blueprint, and perhaps a forewarning of what national censorship regimes could look like in many other countries that have similarly diverse ISP ecosystems to Russia's. As more countries require ISPs to deploy DPI infrastructure for purposes of copyright enforcement or filtering pornography, we risk a slippery slope where Russian-style censorship could easily be deployed. We hope our study is the first in a long line of research into the exact machinations and implications of decentralized control. Such work may be the only way to protect the free and open Internet.

Acknowledgments

The authors are grateful to Patrick Traynor for shepherding the paper, to David Fifield and the anonymous reviewers for their constructive feedback, to Zachary Weinberg and Mahmood Sharif for sharing their knowledge about topic modeling and to our anonymous participants that consented to running our measurements from their network. This work was made possible by the National Science Foundation grant CNS-1755841, and by a Google Faculty Research Award.

References

[1] G. Aceto and A. Pescapé, "Internet censorship detection: A survey," *Computer Networks*, vol. 83, 2015.

[2] Y. Akdeniz, "Internet content regulation: UK government and the control of Internet content," *Computer Law & Security Review*, 2001.

[3] Anonymous, "Towards a Comprehensive Picture of the Great Firewall's DNS Censorship," in *4th USENIX Workshop on Free and Open Communications on the Internet (FOCI 14)*, 2014.

[4] S. Aryan, H. Aryan, and J. A. Halderman, "Internet Censorship in Iran: A First Look," in *FOCI*, 2013.

[5] "Beautiful soup documentation," https://www.crummy.com/software/BeautifulSoup/bs4/doc/.

[6] D. M. Blei, A. Y. Ng, M. I. Jordan, and J. Lafferty, "Latent dirichlet allocation," *Journal of Machine Learning Research*, 2003.

[7] "Что делать если сайт заблокирован провайдером," http://blogsisadmina.ru/internet/chto-delat-esli-sajt-zablokirovan-provajderom.html, November 2015.

[8] A. Bracci and L. Petronio, "New research shows that, post net neutrality, internet providers are slowing down your streaming," https://news.northeastern.edu/2018/09/10/new-research-shows-your-internet-provider-is-in-control/.

[9] "Censored Planet," https://censoredplanet.org/.

[10] A. Chaabane, T. Chen, M. Cunche, E. De Cristofaro, A. Friedman, and M. A. Kaafar, "Censorship in the wild: Analyzing Internet filtering in Syria," in *Proceedings of the 2014 Conference on Internet Measurement Conference*, 2014.

[11] "CDN providers blocked by China," https://www.cdnfinder.com/cdn-providers-blocked-china, 2014.

[12] R. Clayton, S. J. Murdoch, and R. N. Watson, "Ignoring the great firewall of China," in *International Workshop on Privacy Enhancing Technologies*, 2006.

[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.

[14] A. L. Dahir, "Internet shutdowns are costing African governments more than we thought," https://qz.com/1089749/internet-shutdowns-are-increasingly-taking-a-toll-on-africas-economies/, 2017.

[15] S. Darbinyan and S. Hovyadinov, "World Intermediary Liability Map: Russia," https://wilmap.law.stanford.edu/country/russia.

[16] D. Dittrich and E. Kenneally, "The Menlo Report: Ethical principles guiding information and communication technology research," U.S. Department of Homeland Security, Tech. Rep., 2012.

[17] 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.

[18] Z. Durumeric, F. Li, J. Kasten, J. Amann, J. Beekman, M. Payer, N. Weaver, D. Adrian, V. Paxson, M. Bailey *et al.*, "The matter of heartbleed," in *Proceedings of the 2014 conference on internet measurement conference*, 2014.

[19] Z. Durumeric, E. Wustrow, and J. A. Halderman, "ZMap: Fast Internet-wide Scanning and Its Security Applications." in *USENIX Security Symposium*, vol. 8, 2013.

[20] "The art of concealment," https://www.economist.com/special-report/2013/04/06/the-art-of-concealment.

[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.

[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] D. Fifield, C. Lan, R. Hynes, P. Wegmann, and V. Paxson, "Blocking-resistant communication through domain fronting," *PoPETs*.

[24] A. Filasto and J. Appelbaum, "OONI: Open Observatory of Network Interference," in *FOCI*, 2012.

[25] FortiNet, "Fortiguard labs web filter," https://fortiguard.com/webfilter.

[26] S. Frolov, F. Douglas, W. Scott, A. McDonald, B. VanderSloot, R. Hynes, A. Kruger, M. Kallitsis, D. G. Robinson, S. Schultze *et al.*, "An ISP-scale deployment of TapDance," in *7th USENIX Workshop on Free and Open Communications on the Internet (FOCI 17)*, 2017.

[27] G. Gebhart and T. Kohno, "Internet Censorship in Thailand: User Practices and Potential Threats," in *Security and Privacy (EuroS&P), 2017 IEEE European Symposium on*, 2017.

[28] "Gensim Library — gensim 3.8.1," https://pypi.org/project/gensim/.

[29] P. Gill, M. Crete-Nishihata, J. Dalek, S. Goldberg, A. Senft, and G. Wiseman, "Characterizing web censorship worldwide: Another look at the opennet initiative data," *ACM Transactions on the Web (TWEB)*, vol. 9, no. 1, 2015.

[30] "Glasnost: Results from tests for bittorrent traffic shaping," https://broadband.mpi-sws.org/transparency/results/.

[31] "We monitor and challenge internet censorship in China," https://en.greatfire.org.

[32] A. Houmansadr, G. T. Nguyen, M. Caesar, and N. Borisov, "Cirripede: Circumvention infrastructure using router redirection with plausible deniability," in *Proceedings of the 18th ACM conference on Computer and communications security*. ACM, 2011.

[33] A. Hounsel, P. Mittal, and N. Feamster, "Automatically Generating a Large, Culture-Specific Blocklist for China," in *8th USENIX Workshop on Free and Open Communications on the Internet (FOCI 18)*. USENIX Association, 2018.

[34] "Indonesia introduces new internet censorship system," https://www.arabnews.com/node/1218011/world.

[35] "Internet usage statistics," https://www.internetworldstats.com/stats.htm.

[36] "Internet Outage Detection and Analysis, CAIDA," https://ioda.caida.org/ioda/dashboard.

[37] B. Jones, T.-W. Lee, N. Feamster, and P. Gill, "Automated detection and fingerprinting of censorship block pages," in *Internet Measurement Conference (IMC)*. ACM, 2014.

[38] J. Karlin, D. Ellard, A. W. Jackson, C. E. Jones, G. Lauer, D. Mankins, and W. T. Strayer, "Decoy routing: Toward unblockable internet communication." in *FOCI*, 2011.

[39] S. Khattak, M. Javed, P. D. Anderson, and V. Paxson, "Towards Illuminating a Censorship Monitor's Model to Facilitate Evasion," in *FOCI*, 2013.

[40] C. Lab and Others, "Url testing lists intended for discovering website censorship," 2014, https://github.com/citizenlab/test-lists. [Online]. Available: https://github.com/citizenlab/test-lists

[41] "Langdetect Library — langdetect 1.0.7," https://pypi.org/project/langdetect/.

[42] G. Lowe, P. Winters, and M. L. Marcus, "The great DNS wall of China," *MS, New York University*, vol. 21, 2007.

[43] M. Marcon, M. Dischinger, K. P. Gummadi, and A. Vahdat, "The local and global effects of traffic shaping in the internet," *2011 Third International Conference on Communication Systems and Networks (COMSNETS 2011)*, 2011.

[44] B. Marczak, N. Weaver, J. Dalek, R. Ensafi, D. Fifield, S. McKune, A. Rey, J. Scott-Railton, R. Deibert, and V. Paxson, "An Analysis of China's Great Cannon," in *Free and Open Communications on the Internet*. USENIX, 2015.

[45] N. Marechal, "From Russia With Crypto: A Political History of Telegram," in *8th USENIX Workshop on Free and Open Communications on the Internet (FOCI 18)*. USENIX Association, 2018.

[46] "MaxMind," https://www.maxmind.com/.

[47] 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 *Proceedings of the Internet Measurement Conference*, 2018.

[48] D. Meyer, "VPN providers pull Russian servers as Putin's ban threatens to bite," https://www.zdnet.com/article/vpn-providers-pull-russian-servers-as-putins-ban-threatens-to-bite/.

[49] A. Molavi Kakhki, A. Razaghpanah, R. Golani, D. Choffnes, P. Gill, and A. Mislove, "Identifying traffic differentiation on cellular data networks," in *Proceedings of the 2014 ACM Conference on SIGCOMM*, ser. SIGCOMM '14, 2014.

[50] Z. Nabi, "The Anatomy of Web Censorship in Pakistan." in *FOCI*, 2013.

[51] National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, *The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research*, 1978.

[52] "Natural Language Toolkit," https://www.nltk.org/.

[53] "Net Neutrality in the United States," https://en.wikipedia.org/wiki/Net_neutrality_in_the_United_States.

[54] A. Nisar, A. Kashaf, I. A. Qazi, and Z. A. Uzmi, "Incentivizing censorship measurements via circumvention," in *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication, SIGCOMM 2018, Budapest, Hungary, August 20-25, 2018*.

[55] "Open Observatory of Network Interference," https://ooni.torproject.org/.

[56] OONI, "The test list methodology," https://ooni.torproject.org/get-involved/contribute-test-lists/.

[57] 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.

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

[59] N. Perlroth and D. Sanger, "North Korea Loses Its Link to the Internet," https://www.nytimes.com/2014/12/23/world/asia/attack-is-suspected-as-north-korean-internet-collapses.html?_r=0.

[60] "Portuguese ISPs given 40 days to comply with EU net neutrality rules," https://edri.org/portuguese-isps-given-40-days-to-comply-with-eu-net-neutrality-rules/.

[61] "List of websites/domains blocked by ISP's in Portugal," 2019, https://tofran.github.io/PortugalWebBlocking/.

[62] "Lawful interception: the Russian approach, Privacy International," https://www.privacyinternational.org/blog/1296/lawful-interception-russian-approach.

[63] "Register of Internet Addresses Filtered in Russian Federation," https://github.com/zapret-info/z-i.

[64] "Registry of Banned Sites," https://blocklist.rkn.gov.ru.

[65] S. Robertson, "Understanding inverse document frequency: on theoretical arguments for IDF," *Journal of Documentation*, 2004.

[66] "Roskomnadzor," https://rkn.gov.ru/.

[67] "On recommendations of Roskomnadzor to telecom operators on blocking illegal information on the Internet," July 2017, https://archive.fo/LGszb;https://archive.fo/XAgJk.

[68] "University of Oregon Route Views Project," http://www.routeviews.org/.

[69] Rozkomnadzor, "Требования к размещаемой информации об ограничении доступа к информационным ресурса," https://eais.rkn.gov.ru/docs/requirements.pdf, translated from Russian.

[70] 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.

[71] "Snowball," https://snowballstem.org/.

[72] A. Soldatov, "Russia: We know what you blocked this summer," https://www.indexoncensorship.org/2013/10/russia-censored-summer-2013/.

[73] A. Soldatov and I. Borogan, "In Ex-Soviet States, Russian Spy Tech Still Watches You | WIRED," https://www.wired.com/2012/12/russias-hand/.

[74] Soldatov, Andrei, "Russian Surveillance State," https://media.ccc.de/v/29c3-5402-en-russias_surveillance_state_h264.

[75] "Web blocking in the United Kingdom," https://en.wikipedia.org/wiki/Web_blocking_in_the_United_Kingdom.

[76] D. S. L. Ukraine, "Crimea has a website ban list additional to russia-wide list, a research proves," https://medium.com/@cyberlabukraine/crimea-has-a-website-ban-list-additional-to-russia-wide-list-a-research-proves-4f20fa6fc762, September 2018.

[77] J. van der Ham, "Ethics and Internet measurements," in *2017 IEEE Security and Privacy Workshops (SPW)*. IEEE, 2017.

[78] 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.

[79] VASExperts, "DPI для СОРМ, готовимся экономить," https://vasexperts.ru/blog/dpi-dlya-sorm-gotovimsya-ekonomit/.

[80] V. Weber, "The worldwide web of chinese and russian information controls,," https://www.opentech.fund/documents/12/English_Weber_WWW_of_Information_Controls_Final.pdf.

[81] Z. Weinberg, M. Sharif, J. Szurdi, and N. Christin, "Topics of controversy: An empirical analysis of web censorship lists," *Proceedings on Privacy Enhancing Technologies*, 2017.

[82] "What's Happened Since Russia Banned Telegram," https://slate.com/technology/2018/04/russian-internet-in-chaos-because-of-telegram-app-ban.htm/l.

[83] C. Williams, "How Egypt shut down the internet," https://www.telegraph.co.uk/news/worldnews/africaandindianocean/egypt/8288163/How-Egypt-shut-down-the-internet.html, 2011.

[84] P. Winter and S. Lindskog, "How the Great Firewall of China is blocking Tor," 2012.

[85] P. Wintour, "UK ISPs to introduce jihadi and terror content reporting button," https://www.theguardian.com/technology/2014/nov/14/uk-isps-to-introduce-jihadi-and-terror-content-reporting-button.

[86] E. Wustrow, C. M. Swanson, and J. A. Halderman, "Tapdance: End-to-middle anticensorship without flow blocking," in *23rd {USENIX} Security Symposium*, 2014.

[87] E. Wustrow, S. Wolchok, I. Goldberg, and J. A. Halderman, "Telex: Anticensorship in the network infrastructure." in *USENIX Security Symposium*, 2011.

[88] Xu, Xueyang and Mao, Z. Morley and Halderman, J. Alex, "Internet censorship in china: Where does the filtering occur?" in *"Passive and Active Measurement"*, 2011.

[89] 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*, 2018.

[90] "ZDNS," https://github.com/zmap/zdns.

[91] B. Zevenbergen, B. Mittelstadt, C. Véliz, C. Detweiler, C. Cath, J. Savulescu, and M. Whittaker, "Philosophy meets Internet engineering: Ethics in networked systems research," in *(GTC Workshop Outcomes Paper) (September 29, 2015)*.

[92] "ZGrab," https://github.com/zmap/zgrab.

[93] J. Zittrain and B. Edelman, "Internet filtering in China," *IEEE Internet Computing*, 2003.

APPENDIX

A. Validating the Russian Blocklist

To validate our source of historical blocklists at [63], we obtained access to a small set of blocklists digitally signed by Roskomnadzor through a few different anonymous sources. To get the corresponding historical blocklists from the Zapret source, we searched it for the date and timestamp closest to that in the anonymously-supplied blocklists. None of the date and timestamps were a perfect match between the two sources, leading us to believe that the Zapret information has a different source than the small set of blocklists we obtained from our anonymous sources.

Using the closest version of the Zapret source, we preprocessed the contents of both blocklists. This included extracting all IP addresses and all domains in each of the files, resulting in sets of IP addresses and domains to compare between the two different sets of blocklists.

We compared the two sources' sets of IP addresses and domains using the Jaccard index of similarity, which is calculated by taking the size of the intersection of the two sets and dividing by the size of the union of the two sets. The Jaccard index is a number between 0.0 and 1.0, where 0.0 represents no similarity and 1.0 represents completely similar sets.

Date	IPs Only	Domains Only	IPs & Domains
2017-06-13	1.0	0.99998	0.99999
2018-04-27	0.99994	0.99867	0.99805
2018-05-13	0.99733	0.99998	0.99996
2018-11-08	0.99996	0.99999	0.99997

Table VI: **Zapret-supplied blacklists' similarity to anonymously-supplied blacklists signed by Roskomnadzor, using the Jaccard index for each category as mentioned in the column name.** ◇

Applying the Jaccard index to our blocklists from different sources (which was signed by Roskomnadzor), we found that the Zapret blocklists are *extremely* similar to the signed blocklists. Our results are shown in Table VI. We analyzed the similarity of the sets of IP addresses, domains, and the entire set of all IP addresses and domains combined. All sampled blocklists have a similarity greater than 0.99 for any given content type (IP, domain, or IP & domain). Based on these findings, we conclude that the Zapret source of blocklists is representative of the list produced by Roskomnadzor, and is both correct and complete, and thus sufficient for our analysis in the paper.

B. Analysis of RUBL

Figure 8a shows how the number of unique IPs added per day outpaces the number of unique IPs removed per day, further displaying the rapid growth of *RUBL*. In addition, the time series plot shows significant volatility. Significant events such as court rulings restricting a certain service will lead to a spike in number of IPs added. On the other hand, media traction of specific collateral damage instances will lead to a spike in number of IPs removed. Regardless, days without any significant activity is prevalent, leading to many downturns in the graph. Since the blacklisting of specific sites require no court ruling, IP address additions rarely fall to zero. The opposite is true for address removals, which often fall to zero due to the degree of time and difficulty involved for content owners to initiate an official removal procedure. Similar to Figure 2, Figure 8a also shows a rise in addition of unique IPs and decrease in removal of unique IPs in 2019, suggesting that the blocklist is being handled more carefully recently. We observe the same trends with domains in Figure 8b, although there is more variability.

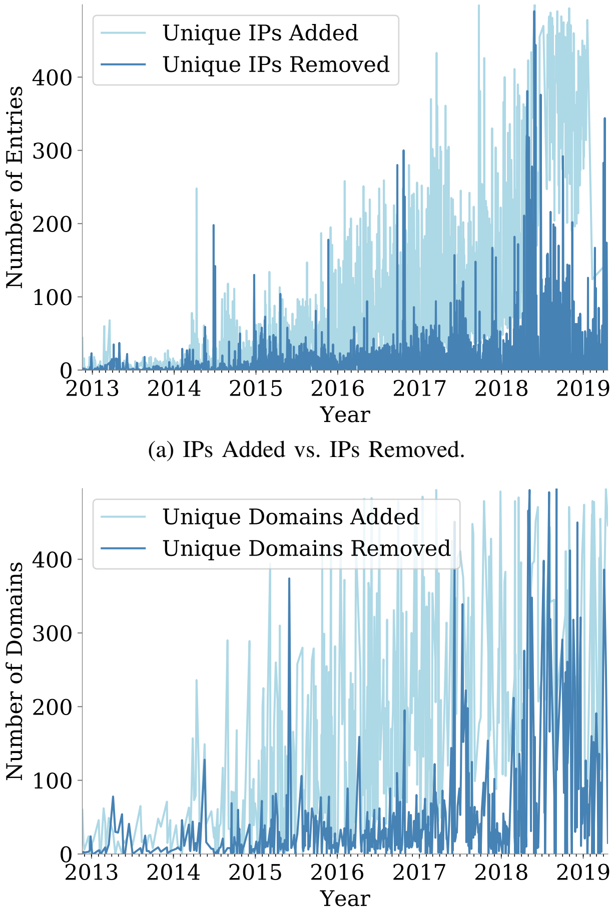
C. Blockpages observed through DNS Poisoning

Figure 10 shows three block pages received at Probe 9, Probe 14, and VPS 6 due to DNS poisoning, as discussed in Section VI-B.

D. $RUBL_{sub}$ Measurements

As we mentioned in Section V-A, the $RUBL_{sub}$ consists of 39 subnets, ranging from /16s to /24s. 31 out of 39 of these subnets contain at least one IP reachable to one of our controls. Of the remaining eight subnets completely unreachable from our controls, seven belong to Telegram and all eight are geolocated to Moscow.

Figure 9 shows the percentage of blocking in each of the 31 subnets in $RUBL_{sub}$ that were reachable in our controls, where percentage of blocking is the number of IPs unreachable out of total number of IPs in the reachable subnets. Two subnets, Subnet 16 and 27, see much lower rates of blocking at most of our probes. These subnets belong to Cloud South, a U.S.-based hosting provider, and UK2, a UK-based hosting provider. Several of the other subnets belong to providers such as DigitalOcean, so it is unclear why these two subnets see less residential blocking, though it might pertain to collateral damage associated with blocking them. Another interesting feature of this analysis is that blocking appears to be correlated with the size of the subnet: larger subnets are blocked more, by both probes and VPSes.



(a) IPs Added vs. IPs Removed.

(b) Domains Added vs. Domains Removed.

Figure 8: **Blocklist volatility over 7 years**—The two subfigures shows the volatility of the blocklist, with many spikes and downturns in response to real world events. ◇

E. Consent form

The consent form used in this study is shown in Figure 11

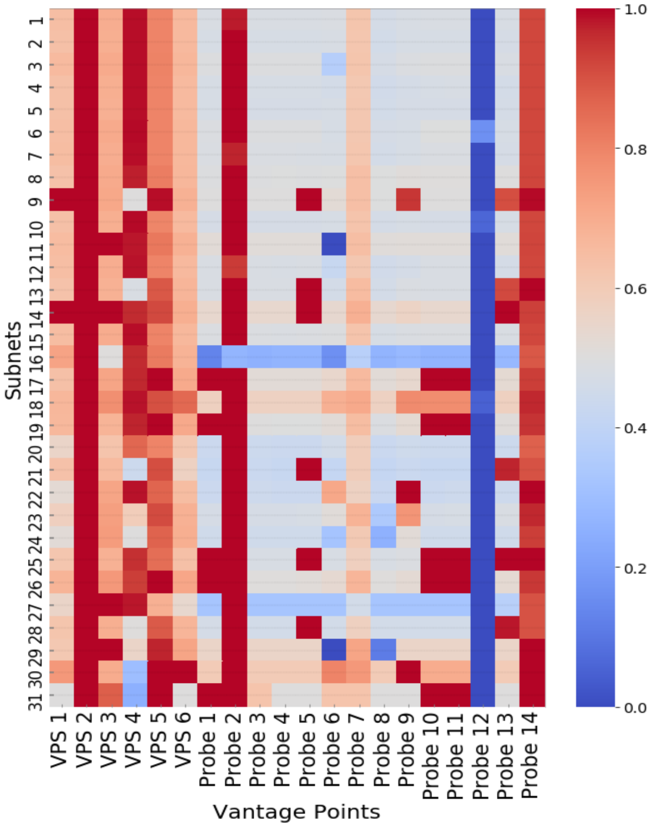
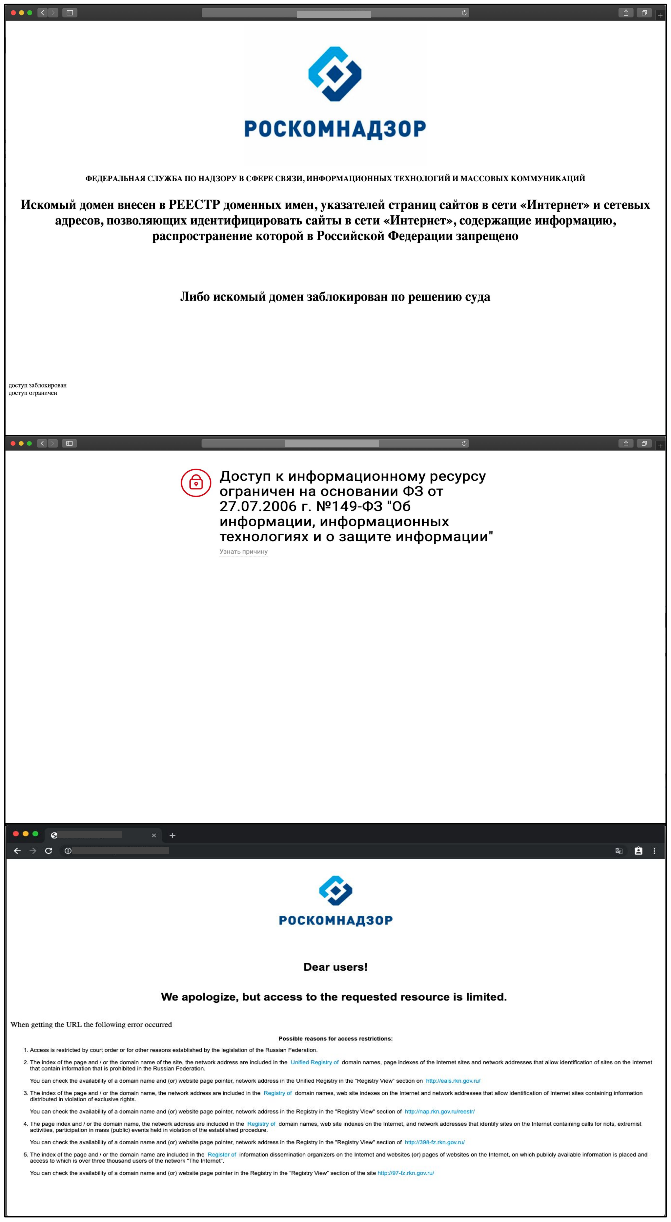


Figure 9: **Blocking per subnet when testing** $RUBL_{sub}$ **on VPSes and Probes**—Datacenter vantage points observe a large percentage of blocking in almost all subnets. Residential vantage point comparatively block intensively in fewer subnets. ◇



ФЕДЕРАЛЬНАЯ СЛУЖБА ПО НАДЗОРУ В СФЕРЕ СВЯЗИ, ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И МАССОВЫХ КОММУНИКАЦИЙ

Искомый домен внесен в РЕЕСТР доменных имен, указателей страниц сайтов в сети «Интернет» и сетевых адресов, позволяющих идентифицировать сайты в сети «Интернет», содержащие информацию, распространение которой в Российской Федерации запрещено

Либо искомый домен заблокирован по решению суда

доступ заблокирован
доступ ограничен

Доступ к информационному ресурсу ограничен на основании ФЗ от 27.07.2006 г. №149-ФЗ "Об информации, информационных технологиях и о защите информации"

Узнать причину

Dear users!

We apologize, but access to the requested resource is limited.

When getting the URL the following error occurred

Possible reasons for access restrictions:

1. Access is restricted by court order or for other reasons established by the legislation of the Russian Federation.
2. The index of the page and / or the domain name of the site, the network address are included in the Unified Registry of domain names, page indexes of the Internet sites and network addresses that allow identification of sites on the Internet that contain information that is prohibited in the Russian Federation.

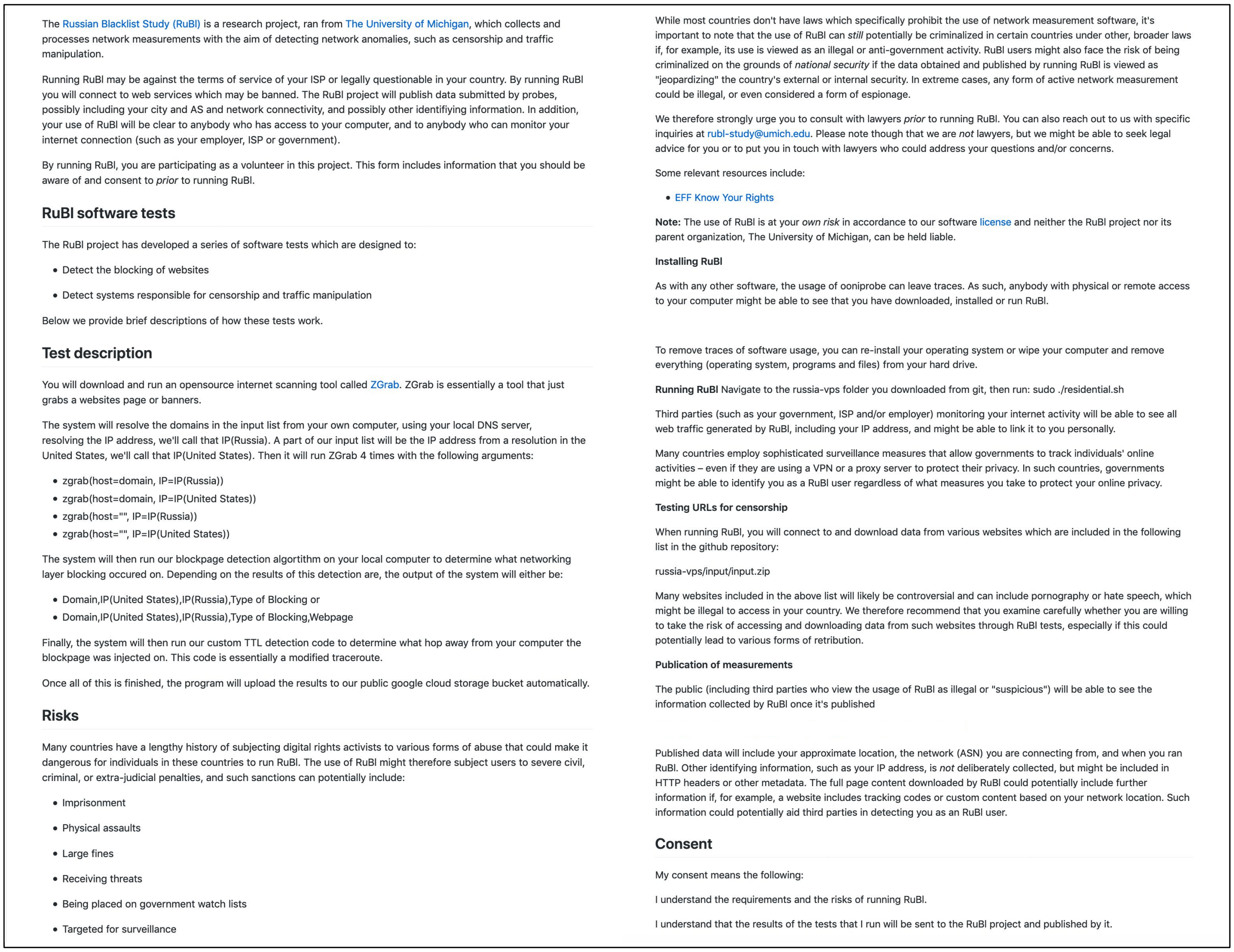
   You can check the availability of a domain name and (or) website page pointer, network address in the Unified Registry in the "Registry View" section on http://eais.rkn.gov.ru/
3. The index of the page and / or the domain name, the network address are included in the Registry of domain names, web site indexes on the Internet and network addresses that allow identification of Internet sites containing information distributed in violation of exclusive rights.

   You can check the availability of a domain name and (or) website page pointer, network address in the Registry in the "Registry View" section of http://nap.rkn.gov.ru/reestr/
4. The page index and / or the domain name, the network address are included in the Registry of domain names, web site indexes on the Internet, and network addresses that identify sites on the Internet containing calls for riots, extremist activities, participation in mass (public) events held in violation of the established procedure.

   You can check the availability of a domain name and (or) website page pointer, network address in the Registry in the "Registry View" section of http://398-fz.rkn.gov.ru/
5. The index of the page and / or the domain name are included in the Register of information dissemination organizers on the Internet and websites (or) pages of websites on the Internet, on which publicly available information is placed and access to which is over three thousand users of the network "The Internet".

   You can check the availability of a domain name and (or) website page pointer in the Registry in the "Registry View" section of the site http://97-fz.rkn.gov.ru/

Figure 10: **Three example blockpages.** ◇



The Russian Blacklist Study (RuBl) is a research project, ran from The University of Michigan, which collects and processes network measurements with the aim of detecting network anomalies, such as censorship and traffic manipulation.

Running RuBl may be against the terms of service of your ISP or legally questionable in your country. By running RuBl you will connect to web services which may be banned. The RuBl project will publish data submitted by probes, possibly including your city and AS and network connectivity, and possibly other identifiying information. In addition, your use of RuBl will be clear to anybody who has access to your computer, and to anybody who can monitor your internet connection (such as your employer, ISP or government).

By running RuBl, you are participating as a volunteer in this project. This form includes information that you should be aware of and consent to *prior* to running RuBl.

RuBl software tests

The RuBl project has developed a series of software tests which are designed to:

- Detect the blocking of websites
- Detect systems responsible for censorship and traffic manipulation

Below we provide brief descriptions of how these tests work.

Test description

You will download and run an opensource internet scanning tool called ZGrab. ZGrab is essentially a tool that just grabs a websites page or banners.

The system will resolve the domains in the input list from your own computer, using your local DNS server, resolving the IP address, we'll call that IP(Russia). A part of our input list will be the IP address from a resolution in the United States, we'll call that IP(United States). Then it will run ZGrab 4 times with the following arguments:

- zgrab(host=domain, IP=IP(Russia))
- zgrab(host=domain, IP=IP(United States))
- zgrab(host="", IP=IP(Russia))
- zgrab(host="", IP=IP(United States))

The system will then run our blockpage detection algortithm on your local computer to determine what networking layer blocking occured on. Depending on the results of this detection are, the output of the system will either be:

- Domain,IP(United States),IP(Russia),Type of Blocking or
- Domain,IP(United States),IP(Russia),Type of Blocking,Webpage

Finally, the system will then run our custom TTL detection code to determine what hop away from your computer the blockpage was injected on. This code is essentially a modified traceroute.

Once all of this is finished, the program will upload the results to our public google cloud storage bucket automatically.

Risks

Many countries have a lengthy history of subjecting digital rights activists to various forms of abuse that could make it dangerous for individuals in these countries to run RuBl. The use of RuBl might therefore subject users to severe civil, criminal, or extra-judicial penalties, and such sanctions can potentially include:

- Imprisonment
- Physical assaults
- Large fines
- Receiving threats
- Being placed on government watch lists
- Targeted for surveillance

While most countries don't have laws which specifically prohibit the use of network measurement software, it's important to note that the use of RuBl can *still* potentially be criminalized in certain countries under other, broader laws if, for example, its use is viewed as an illegal or anti-government activity. RuBl users might also face the risk of being criminalized on the grounds of *national security* if the data obtained and published by running RuBl is viewed as "jeopardizing" the country's external or internal security. In extreme cases, any form of active network measurement could be illegal, or even considered a form of espionage.

We therefore strongly urge you to consult with lawyers *prior* to running RuBl. You can also reach out to us with specific inquiries at rubl-study@umich.edu. Please note though that we are *not* lawyers, but we might be able to seek legal advice for you or to put you in touch with lawyers who could address your questions and/or concerns.

Some relevant resources include:

- EFF Know Your Rights

Note: The use of RuBl is at your *own risk* in accordance to our software license and neither the RuBl project nor its parent organization, The University of Michigan, can be held liable.

Installing RuBl

As with any other software, the usage of ooniprobe can leave traces. As such, anybody with physical or remote access to your computer might be able to see that you have downloaded, installed or run RuBl.

To remove traces of software usage, you can re-install your operating system or wipe your computer and remove everything (operating system, programs and files) from your hard drive.

Running RuBl Navigate to the russia-vps folder you downloaded from git, then run: sudo ./residential.sh

Third parties (such as your government, ISP and/or employer) monitoring your internet activity will be able to see all web traffic generated by RuBl, including your IP address, and might be able to link it to you personally.

Many countries employ sophisticated surveillance measures that allow governments to track individuals' online activities – even if they are using a VPN or a proxy server to protect their privacy. In such countries, governments might be able to identify you as a RuBl user regardless of what measures you take to protect your online privacy.

Testing URLs for censorship

When running RuBl, you will connect to and download data from various websites which are included in the following list in the github repository:

russia-vps/input/input.zip

Many websites included in the above list will likely be controversial and can include pornography or hate speech, which might be illegal to access in your country. We therefore recommend that you examine carefully whether you are willing to take the risk of accessing and downloading data from such websites through RuBl tests, especially if this could potentially lead to various forms of retribution.

Publication of measurements

The public (including third parties who view the usage of RuBl as illegal or "suspicious") will be able to see the information collected by RuBl once it's published

Published data will include your approximate location, the network (ASN) you are connecting from, and when you ran RuBl. Other identifying information, such as your IP address, is *not* deliberately collected, but might be included in HTTP headers or other metadata. The full page content downloaded by RuBl could potentially include further information if, for example, a website includes tracking codes or custom content based on your network location. Such information could potentially aid third parties in detecting you as an RuBl user.

Consent

My consent means the following:

I understand the requirements and the risks of running RuBl.

I understand that the results of the tests that I run will be sent to the RuBl project and published by it.

Figure 11: **The Consent Form.** ◇