A Bestiary of Blocking
The Motivations and Modes behind Website Unavailability

Michael Carl Tschantz
ICSI

Sadia Afroz
ICSI and UC Berkeley

Shaarif Sajid
LUMS

Shoaib Asif Qazi
LUMS

Mobin Javed
LUMS

Vern Paxson
ICSI and UC Berkeley

Abstract

This paper examines different reasons that websites may vary in their availability by location. Prior works on availability mostly focus on censorship by nation states. We measure three forms of server-side blocking: blocking visitors from the EU to avoid GDPR compliance, blocking based upon the visitor's country, and blocking due to security concerns. We argue that these and other forms of blocking warrant more research.

1 Introduction

We often conceptualize the Internet as one global and shared infrastructure comprehensively connecting people from every part of the world. In practice, however, different users experience different Internets. The differences in experience can arise for various reasons, such as ISPs creating restricted "walled gardens" for their customers, governments censoring access to resources, copyright regulations restricting access to protected content, and web servers blocking unwanted access. These partionings of the Internet in terms of the way content is served to different end users reflect instances of the "balkanization"[1] of the Internet into a "splinternet".

Currently, a large body of research exists on understanding access restrictions by authoritarian states for censorship (e.g., [50, 14, 32, 17, 38, 41, 13, 5, 44, 46, 30, 37, 4, 18, 3, 11, 33, 19]). The quintessential case of censorship is a government blocking communication between two willing parties to further political control. That is, censorship is typically seen as involving a particular type of entity doing the blocking, a government, and with a particular type of motivation, the exercise of power to support a political agenda. Exactly delimiting which parties count as government actors is complicated by the possibility of the government indirectly acting through pressuring others to do its bidding. Precisely defining *political control* is even more difficult.

Even without resolving these definitional difficulties, we can conclude that many of the aforementioned forces leading to regional differences in the Internet do not fit under many reasonable conceptions of censorship. For example, some of these forms of access restriction are controlled by algorithms running on a website's servers or CDNs, not the government, for the websites' own purposes, such as profit maximization, not political control. These forms of restriction have received much less empirical exploration than censorship and may be mistaken for censorship by naive measurement methods.

Herein, we explore forms of that blocking that do not fit into the quintessential conceptualization of censorship—we leave it to the reader to decide which count as censorship in the broadest sense of the term. We start by enumerating various types of blocking and considering how they relate to censorship (Section 2).

Next, we present measurements for three such forms of blocking. For each measurement, we loaded webpages from various locations, and when a page would only properly load in some of the locations, we examined error codes and block pages, if any, to see whether they provided a reason (Section 3). We look for, and find, blockpages explicitly listing geography as the reason for blocking (Section 4). We also look at blockpages and practices suggesting security concerns (Section 5). Finally, we show that the number of pages unavailable from three locations in the EU increased after the EU's General Data Protection Regulation (GDPR) went into effect (Section 6).

In each of these cases, it appears that the blocking is done by the website's owners, and not the government. While we cannot rule out a government indirectly causing these webpages to be blocked by putting pressure upon the website owners, this seems unlikely for the blocking that suggest security concerns. The case of blocking to avoid GDPR compliance is more debatable.

[1] For a critical discussion of the term see Maurer [34].

This blocking is a workaround in response to a government mandate, but not motivated by the sort of political control associated with quintessential censorship. In fact, the website owners could comply with the regulation without changing the webpages' content.

While many of these issues have been discussed, and in some cases measured, in isolation, we believe this work to be the first to consider the range of blocking in a systematic manner. We also discuss the difficulties of separating out each form from the others, which is further complicated by the possibility of a single block corresponding to more than one form. Our contributions are, admittedly preliminary: our list of blocking types is incomplete and tilted toward location-based blocking; our measurement studies are small-scale. Nevertheless, we believe there to be value in laying out this space of research opportunities while highlighting the risks of claiming to measure only a single phenomenon given the lack of isolation between the types of blocking we consider.

Prior Work. Prior work has not presented the space of blocking forms, which is the goal of this work. However, there are numerous papers looking at various forms of blocking in isolation. Thus, rather than have a dedicated prior work section, we will cover these works where we discuss the form of blocking they cover, mostly in the next section.

2 Types of Blocking

Suppose you run a test to find that a website will load in the US but not in China. If the website is politically sensitive, it is not unreasonable to suspect censorship, but numerous other possibilities exist.

Perhaps the first thing to check is the nature of the block: was it just a transient network failure? did DNS fail? is there a CAPTCHA? a blockpage providing an explanation? a blockpage without an explanation? or an error message? For example, while CAPTCHAs can be annoying, they seem like an unlikely choice for censorship since they, when working as designed, merely slow down the accessing of data. Each of the others seem like stronger indications of censorship, including, to a lesser extent, even blockpages claiming the cause to be something else, since censors may mislead.

Another factor to check is what the blocking is targeting. The blocking could be targeting something other location, such as the OS or browser used, automated bots loading pages, or being logged out of a service. For example, some websites block Tor [29, 42]. Keeping these factors and other factors consistent across the two locations can help rule them out, leading to location-based blocking becoming the most likely explanation.

Questions will still remain about what sort of location is targeted by the block. The targeted location might be geographic, such as a campus, sub-national region, country, or super-national region. Alternatively, the targeted location might be defined in terms of network topology: an IP address, an IP address range, a network, or an AS. One can also ask whether the blocking is whitelisting or blacklisting. In *whitelisting*, a website aims to serve its content to only visitors within its region. In *blacklisting*, a website aims to exclude certain regions. Censorship could target any of these notions of location, but the blacklisting of a whole country (the government's own country) is the most characteristic of censorship.

Determining the target of the block and its mode of operation can be tricky. For example, a large enough blacklist will look like a whitelist, and blocking enough IP addresses individually will eventually block a whole range or even a whole country. Furthermore, given that geographic blocks are likely implemented by blocking IP address ranges assigned to the targeted country, targeting can be considered at multiple levels from specification to implementation. Nevertheless, in Section 4, we are able to find country-based blocking with high confidence by finding blockpages that confess to it.

Another factor is where in the network the blocking is happening. The paragon of censorship is a government-operated middlebox at the national border. However, other possibilities exist. The ISP of the client might be doing the blocking (e.g., [17]), or the ISP of the server hosting the tested website, or the website itself. A combination of examining how the block operates and additional measurements can sometimes determine which of these possibilities it is [17, 48, 4, 1].

However, even determining where in the network the blocking is happening does not definitively reveal whether censorship is in action. Suppose, you find that the blocking is done by the ISP of your client in China. This could be because the government of China ordered the block or because the ISP is performing some sort of traffic filtering, in violation of net neutrality, to raise more revenue or cut costs.

Alternatively, suppose you find that the website's server is doing the blocking. At first, this might seem to be a clear indication that the block is for some reason other than censorship, such as concerns over abuse. However, this could still be an instance of China censoring the website, just in the more roundabout manner of pressuring the website into blocking visitors from China. Indeed, Western companies have altered their websites for Chinese visitors to comply with China's demands [10]. Alternately, it could be the server's country doing the censorship by ordering the blocking of visitors from China.

In fact, government orders are behind many sorts of server-side blocking that might or might not strike the reader as censorship. The US's economic sanctions cause websites to block countries [6, 23]. Another recent example is the passage of SESTA, a US law holding websites liable for some third-party content facilitating prostitution, which has lead to a website geo-blocking the US [8]. An example we will explore is websites blocking the EU to avoid GDPR compliance (Section 6). Where to draw the line as to which count as censorship is unclear to us, but server-side censorship of one form or another is possible.

With this mind, it is clear that censorship is not merely an issue of where the blocking is happening or who is doing it, but rather also one of why the blocking is happening, that is, upon whose orders. In some cases, the roles might be switched from what is expected. For example, arguably, copyright is a form of censorship in which the copyright holder gets a government to enforce its claim [43, 40, 2] In theory, a website could pay a government to implement a regional block to reduce abuse or increase its market share, leading to an odd form of hybrid government–corporate censorship.

Before concluding that censorship has happened, other possible motivations behind the block should be considered. Table 1 provides a partial list of different forms of blocking. One possibility is security concerns, such as fraud and abuse, which is associated with certain countries [9, 1]. We look at such blocking in Section 5. Another possibility is concerns over the costs of serving traffic to some countries, which can be seen as a wasted expense for companies not targeting that market. This issue can be exacerbated by the serving of traffic to the developing world sometimes being more expensive than serving it to the developed world [1].

Also with profits in mind, companies may engage in blocking to increase revenue, by charging extra fees to access some regions or by blocking competitors, violations of net neutrality [28]. While not blocking, some websites have engaged in price discrimination, which can also negatively affect some visitors based upon their location [35, 36, 45, 24].

Finally, blocks can be unintentional, for example, from misconfiguration or failures caused by lacking enough bandwidth [47, 27, 49].

The numerous forms of blocking we have mentioned are not independent of one another. For example, some serve as implementation approaches for others. Measurement studies must take care not to conflate forms of blocking. The obvious way of doing so is to just ignore the differences. Less obvious is conflation by attempting identify a form using proxies for it without making assumptions explicit, such as assuming that no server-side blocking is censorship. Developing methods for distinguishing between blocking types could also aid the blocked users, who currently struggle to understand what is happening and why [21]

3 Methods

From our prior work [1], we re-used a crawler, and, in some cases, data. The crawler attempts DNS resolution for each provided URL, logging any errors. For those that resolved, it uses Python's Requests package to request the webpage, logging errors, status codes, and content. The crawler uses a timeout of 30 seconds and Chrome user agent string from a MacBook. It attempts to load all pages with HTTP, but follows any automatic switches to HTTPS. If the DNS resolution fails and the URL lacks the "www." prefix, the crawler tries again with it added. See our prior work for details [1].

For each measurement study presented below, we used the crawler to attempt to download a selection of URLs at various times and from various regions around the world, which vary for the study. We look for differences in errors, the status codes, and content from one load attempt to another. Theses differences can indicate a webpage being available in one location but not another, or at one time but not another, based upon what varied between the load attempts.

To identify the reason behind this difference, we focus on websites returning an explanatory blockpage (with either a 200 or non-200 status code) to one location and standard content to another. While this method gives us reasonably high confidence in the cause, we do not have certainty since the blockpage could be misleading or even injected by a middlebox masquerading as the website. Furthermore, this method is limited to cases where the website volunteers a reason.

Future work can attempt to determine the cause when not volunteered or to confirm the truthfulness of blockpages. Our prior work [1] makes a first step in this direction by using traceroutes to rule out spoofed blockpages from masquerading middleboxes.

4 Country-based Blocking

Cloudflare, a CDN, is an interesting subject of study, not only because it hosts many websites, but because it provides more information than many explaining why it blocks, on the behalf of the website owner, certain requests. In this section and the next, we analyze Cloudflare's block notices looking for country-based and then security-motivated blocking. We emphasize that we selected Cloudflare not because we believe it to engage in such blocks any more than any other host, nor because we believe it should be singled out for criticism. Rather,

Table 1: Examples of Motivations behind Location-based Blocking. Those marked with * denote location in the network topology instead of geo-location. ** denotes cases where we use a non-location-based block due to not finding a location-based one.

	Server (including CDNs)	Middlebox (ISPs, governments)
Political censorship	Bowing to China's demands [10]	Great Firewall of China (lots)
Economic sanctions	US websites blocking Cuba [6] & Iran [23]	
Third-party liability	Blocking US due to SESTA [8]	
Copyright	YouTube blocking in Germany [43, 40]	ISPs blocking the Pirate Bay [2]
Other compliance	GDPR (§6)	
Security	Blocking countries assoc. w. fraud [9, 1] (§5)	
Hosting costs	CDN fees [1]	
Revenue	Price discrimination [35, 36, 45, 24]	Net-neutrality [28]*
Unintentional	Slash-dotting [47]**	Overloaded rural links [27, 49]

we selected Cloudflare since prior work has found it blocking Tor based on abuse [29] and because of the information that Cloudflare provides about blocks, a commendable feature.

We start by finding sites hosted by Cloudflare. We resolved the top 1M Alexa domains, and identified those hosted by Cloudflare by performing *whois* lookups on the resolved IP addresses and keeping those containing "Cloudflare" in the AS name. We identified 85,421 Cloudflare-hosted URLs in this fashion.

Next, for each of these Cloudflare-hosted URLs, we retrieved the website from five vantage points: Pakistan (home network), Scotland (VPN), South Africa (cloud), Ukraine (VPN), and the US (institutional network). The US crawler experienced a failure, limiting its collection to 77,935 URLs.

We then classified the responses. By examining the blockpages themselves and Cloudflare's documentation [22, 15], we determined that a response with an HTTP status code of 403 and a body that referenced Cloudflare's own error code 1009 indicates blocking by country. Table 2 shows a breakdown of the responses by both this code and others, some of which will be relevant to the next section.

We found 524 websites using country-based blocking by Cloudflare under error code 1009. We also found one website using country-based blocking by a different service provider, Dell's SonicWall, which the website `motorcar.com` used in addition to Cloudflare. The error message for SonicWall's 403 blockpage says "Sorry this content is not available in your country due to GDPR", shown in Figure 1, proving to be instance of both GDPR-based and country-based blocking.

Interestingly, 32 websites were country blocked in the US, with 21 giving a Cloudflare 1009 error. We manually

Figure 1: SonicWall Blockpage Showing the Motivation to be GDPR avoidance.

checked all 32 and found five that would load, including one that was give a 1009 error to the crawler. A different website with a 1009 error was `aquapro.biz`. It was blocking countries and manually unavailable in the US, but consistently misidentified our country in an inconsistent manner, seemingly to assign us other countries at random. The remaining 1009 errors explicitly blocked the US. (For an example, see Figure 2.)

While the rate of country blocks varied from country to country, this comparison is complicated by the fact that different countries had different success rates at getting any response from the server. For example, Pakistan had an abnormally high rate of DNS errors, possibly due to network failures or censorship. This difference might hide a much higher rate of blocks in Pakistan than in Scotland. Alternatively, Scotland using a VPN and Pakistan using a home network might hide the difference. However, Scotland and Ukraine can be compared on a fairly even basis for both of these factors. For them, we see a large difference with Ukraine receiving more blocks.

Table 2: Blockpage types for 85,421 Cloudflare-hosted domains from various vantage points.

Blocktypes / Vantage point		Ukraine	Scotland	Pakistan	South Africa	USA
Status	Description	(VPN)	(VPN)	(Home)	(Cloud)	(Inst.)
No HTTP Response						
n/a	Timed out	579	542	607	577	540
n/a	DNS error	45	112	4096	4	66
n/a	Other connection errors	147	959	132	70	525
Geo-blocking totals		**313**	**175**	**178**	**103**	**32**
403	Cloudflare: country or region blocked (1009)	257	161	162	88	21
403	SonicWall Geo-IP filter	1	1	0	1	0
403	Other blockpage mentioning geo-blocking	40	11	3	3	0
200	Other blockpage mentioning geo-blocking	15	1	13	11	10
451	Unavailable For Legal Reasons	0	1	0	0	1
Abuse-blocking totals		**3431**	**1417**	**1874**	**1537**	**1255**
403	Cloudflare: IP Blocked (1006, 1007, 1008)	23	5	6	5	1
200	Cloudflare: IP Blocked (1006, 1007, 1008)	2	2	2	2	1
503	Cloudflare: Browser Verification	1519	1091	1111	1124	985
200	Cloudflare: Browser Verification	2	2	3	2	3
403	Cloudflare: CAPTCHA Challenge	1874	309	746	395	257
403	OctoNet HTTP filter: VPN / TOR Block	3	0	0	0	0
Misconfigurations totals		**8**	**8**	**6**	**9**	**8**
403	Cloudflare: DNS points to invalid IP (1000, 1002)	8	8	6	9	8

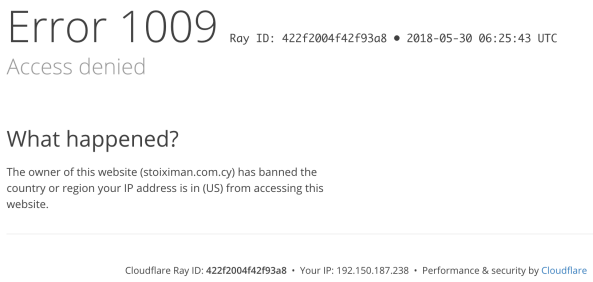

Figure 2: Cloudflare Blockpage with a 1009 Error Code

5 Security-motivated Blocking

Country-based blocking can be viewed as an implementation approach. In this section and the next we look at the motivations behind blocks.

First, using the data set described in Section 4, we looked for security-motivated blocking. While we recognize that some country-based blocks may have security as its motivation, we exclude those country-based blockpages discussed in Section 4 to focus on those not yet discussed. We look at other types of blocking that could have been motivated by security concerns, while admitting that we cannot be sure of the real motivations behind a block.

Again, Cloudflare's documentation helped us know where to look [22, 15]. The most indicative error code of security concerns is 1010, described as "bad browser", which happens when "The source of the request was not legitimate or the request itself was malicious" [22]. Cloudflare also uses error code 1012 to deny access "based on malicious activity detected from your computer or your network (ip_address)" [16]. We also included error codes indicating an IP block, although those could be used for non-security reasons.

Finally, we looked at restrictions short of outright blocking. Namely, Cloudflare will sometimes make users solve a CAPTCHA before showing them the page. Cloudflare will also use a "browser challenge" on visitors it suspects of being a bot, which we looked for (see Figure 3).

Table 2 shows how common each of these, and other, outcomes are. As with country-based blocking, comparing across countries is confounded. Looking again at the well matched pair of Scotland and Ukraine, we see a large difference, with Ukraine receiving more blocks, CAPTCHAs, browser verifications. Between the two, only Ukraine was accused of being a VPN or Tor de-

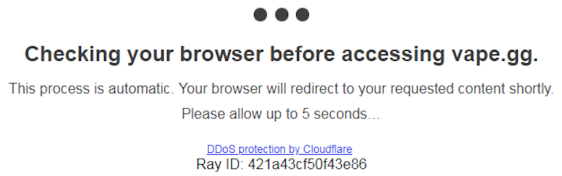

Figure 3: Cloudflare's Browser Challenge in Action

spite both using the same VPN provider. The VPN/Tor blockpages came not from Cloudflare, but from OctoNet HTTP filter.

6 GDPR-Motivated Geo-blocking

The EU's General Data Protection Regulation (GDPR) contains a wide range of provisions designed to protect the privacy of people who use online services and to give them more control over their data [20]. Complying with some of the provisions may require a major shift in how some websites store and process data about their visitors. (See Lomas [31] for an overview.) For example, in general, visitors have the right to access, correct, and delete data about themselves. Implementing these abilities can create an implementational headache given systems engineered to use and store data indiscriminately. Furthermore, getting it correct is high stakes, with fines of €20M or 4% of a company's global annual revenue. Given the uncertainty and stakes, some websites have decided to exit the European market, at least for the time being [25, 26]. To partly quantify this effect, we analyze the differences in availability of a convenience sample of websites before and after GDPR went into effect.

From our prior work [1], we re-used a data set showing the availability of 7081 websites, which we collected to study a different facet of server-side blocking. These websites form the union of various Alexa top 500 lists: the global list, the lists for ten countries, and the lists for nine categories of websites. For each URL, we measured it from three locations in the EU via a VPN: London, United Kingdom; Sofia, Bulgaria; and Frankfurt, Germany. We also use measurements from the US for comparison purposes. From each location, we use one of the measurements of the URLs before GDPR came into effect.

After GDPR came into effect, we re-used the crawler to take a second measurement of each URL. We analyzed the data for changes in website availability.

Using error logs and status codes, we found 74 websites that, for all three European locations, sent an HTTP status code of 200 *OK* when accessed before May 25 and

Table 3: Websites explicitly mentioning GDPR as motivation for blocking. The *Before* column shows the status for all the vantage points. DE represents DNS Error.

	Before	After		
URL/Country		US	BGR	GBR	DEU
bismarcktribune.com	200	200	DE	403	403
collegian.psu.edu	200	200	403	403	403
dailynebraskan.com	200	200	DE	403	403
dailyprogress.com	200	200	403	403	403
fredericknewspost.com	200	200	403	403	403
fredericksburg.com	200	200	403	DE	403
globegazette.com	200	200	403	403	403
greensboro.com	200	200	403	403	403
gwinnettdailypost.com	200	200	403	403	403
havasunews.com	200	200	403	403	403
heraldtimesonline.com	200	200	403	403	403
host.madison.com/wsj	200	200	403	403	403
journalnow.com	200	200	403	403	403
journalstar.com	200	200	403	403	403
journaltimes.com	200	200	DE	403	403
lacrossetribune.com	200	200	403	403	403
lancasteronline.com	200	200	403	403	403
napavalleyregister.com	200	200	DE	403	403
nwitimes.com	200	200	DE	403	403
omaha.com	200	200	403	403	403
pantagraph.com	200	200	403	403	403
pilotonline.com	200	200	403	403	403
postandcourier.com	200	200	403	403	403
postbulletin.com	200	200	403	403	403
pressofatlanticcity.com	200	200	DE	403	403
qctimes.com	200	200	403	403	403
rapidcityjournal.com	200	200	403	403	403
richmond.com	200	200	403	403	403
roanoke.com	200	200	403	403	403
santafenewmexican.com	200	200	403	403	403
southbendtribune.com	200	200	403	403	403
stltoday.com	200	200	403	403	403
theadvocate.com	200	200	403	403	403
trib.com	200	200	403	403	403
tucson.com	200	200	403	403	403
wacotrib.com	200	200	403	403	403
wcfcourier.com	200	200	DE	403	403
wvgazettemail.com	200	200	403	403	403
yakimaherald.com	200	200	403	403	403

Table 4: Websites mentioning "Blocked for legal reasons". The *Before* column shows the status for all the vantage points. DE represents DNS Error.

	Before	After			
URL/Country		US	BGR	GBR	DEU
ctpost.com	200	200	451	451	451
greenwichtime.com	200	200	451	451	451
lmtonline.com	200	200	DE	451	451
newstimes.com	200	200	451	451	451
nhregister.com	200	200	451	451	451
seattlepi.com	200	200	451	451	451
stamfordadvocate.com	200	200	451	451	451

non-200 status after May 25, 2018. Out of the 74 websites, 40 responded with a 403 *Forbidden* status code and a block page explicitly mentioning "Blocked due to GDPR" (Table 3). Seven websites used the HTTP status code 451 *Unavailable For Legal Reasons* (Table 4), the code named for a novel on censorship [7]. All 47 of these websites with explicit blockpages are local news websites, incidentally, a plausible target of censorship as well. One website, `brownells.com`, asks users to visit their EU website, `www.brownells.eu`.

The remaining 27 websites whose availability changed for all three locations do not provide any explicit blockpages and use rather vague status codes and connection errors. For example, the online gaming website `addictinggames.com` returns an empty page with a 404 *Page Not Found* status code, the math tutoring website `webmath.com` refuses the TCP connection, and the reality website `99acres.com` responded with a 412 *Precondition Failed* error code.

Other websites' availability changed for some but not all of the locations. For example, after GDPR, `latimes.com` loaded in Bulgaria, but not Germany and the UK. Its block page states that it is "unavailable in most European countries" as they "identify technical compliance solutions", but does not name GDPR.

Providing context to our findings, looking online, we found services aiming to make it easy to block all EU visitors [39, 12].

7 Conclusions

We have laid out a space of blocking that includes, but also exceeds, what we normally think of as censorship. We looked at three such forms of blocking in some detail. One of them, country-based blocking, is directly tied to location. It is more of an approach for implementing blocking than a motivation for blocking, raising the question of why the blocking is happening. The other two forms we measured are more like motivations than implementation approaches. One of them, avoiding GDPR compliance, is directly related to location in that the websites are blocking visitors from the EU for being from the EU. The other, security-motivated blocks, differs in that it does not need to be implemented using locations. However, we do find large differences in how common security-based blocks are from one location to the next, even when using the same VPN service to send requests from each location. While we studied three forms of blocking, they were far from independent of one another. For example, security concerns might have motivated some country-based blocks. While each of our studies were small scale, we hope they stimulate further research on these issues.

Deciding exactly which of these count as *censorship* is politically fraught, and we will not attempt to do so. We do take the stance that research should cover all forms of blocking that can adversely affect some users, particularly, when those effects are concentrated on people from certain countries.

Furthermore, we believe the chilling effects on website availability of even well-intentioned laws to be an interesting subject of measurement. While we may wish for a world with both the robust privacy protections of the GDPR and an Internet free from balkanization, currently, a tradeoff is evident. The blocking of EU visitors precipitated by a privacy law may even have an outsized effect on Tor given the outsized number of Tor exits in the EU. This serves as an example of how the forms of blocking do not just have interdependencies between themselves but also with privacy. The presence of these interdependencies should be kept in mind when attempting to measure censorship to avoid false positives. The motivations and actors behind blocking are crucial for understanding the blocking, however, identifying them might require additional measurements.

Acknowledges. We thank David Fifield for allowing us to use his code for categorizing block pages, and Narseo Vallina Rodriguez and Mohammad Taha Khan for allowing us to use their paid VPN services. We also thank Jenna Burrell for discussing her experiences of server-side blocking from Ghana, which initiated this project. We gratefully acknowledge funding support from the National Science Foundation (Grants 1518918 and 1651857) and UC Berkeley's Center for Long-Term Cybersecurity. The opinions in this paper are those of the authors and do not necessarily reflect the opinions of any funding sponsor or the United States Government.

References

[1] AFROZ, S., TSCHANTZ, M. C., SAJID, S., QAZI, S. A., JAVED, M., AND PAXSON, V. Exploring server-side blocking of regions. *ArXiv 1805.11606* (May 2018).

[2] ALBANESIUS, C. U.K. high court orders ISPs to block the Pirate Bay. *PC Mag* (Apr. 2012). https://www.pcmag.com/article2/0,2817,2403749,00.asp.

[3] ANONYMOUS. Towards a comprehensive picture of the Great Firewall's DNS censorship. In *Free and Open Communications on the Internet* (2014), USENIX.

[4] ARYAN, S., ARYAN, H., AND HALDERMAN, J. A. Internet censorship in Iran: A first look. *Free and Open Communications on the Internet, Washington, DC, USA* (2013).

[5] BAMMAN, D., O'CONNOR, B., AND SMITH, N. A. Censorship and deletion practices in chinese social media. *First Monday 17*, 3 (Mar. 2012).

[6] BISCHOF, Z. S., RULA, J. P., AND BUSTAMANTE, F. E. In and out of cuba: Characterizing cuba's connectivity. In *Proceedings of the 2015 Internet Measurement Conference* (New York, NY, USA, 2015), IMC '15, ACM, pp. 487–493.

[7] BRADBURY, R. *Fahrenheit 451*. Ballantine Books), 1953.

[8] BRODKIN, J. "erotic review" blocks US Internet users to prepare for government crackdown. *Ars Technica* (Apr. 2018). https://arstechnica.com/tech-policy/2018/04/erotic-review-blocks-us-internet-users-to-prepare-for-government-crackdown/.

[9] BURRELL, J. *Invisible users: Youth in the Internet cafés of urban Ghana*. Mit Press, 2012.

[10] CARSTEN, P. Microsoft denies global censorship of China-related searches. *Reuters* (Feb. 2014). https://www.reuters.com/article/us-microsoft-bing-censorship-idUSBREA1B0CP20140212.

[11] CHAABANE, A., CHEN, T., CUNCHE, M., CRISTOFARO, E. D., FRIEDMAN, A., AND KAAFAR, M. A. Censorship in the wild: Analyzing Internet filtering in Syria. In *Internet Measurement Conference* (2014), ACM.

[12] CIMPANU, C. New service blocks EU users so companies can save thousands on GDPR compliance. *Bleeping Computer* (May 2018). https://www.bleepingcomputer.com/news/security/new-service-blocks-eu-users-so-companies-can-save-thousands-on-gdpr-compliance/.

[13] CITIZEN LAB. Behind Blue Coat: Investigations of commercial filtering in Syria and Burma. https://citizenlab.org/2011/11/behind-blue-coat/, Nov. 2011.

[14] CLAYTON, R., MURDOCH, S. J., AND WATSON, R. N. M. Ignoring the Great Firewall of China. In *Privacy Enhancing Technologies* (Cambridge, England, 2006), Springer, pp. 20–35.

[15] CLOUDFLARE. Error pages, troubleshooting, cloudflare support. https://support.cloudflare.com/hc/en-us/sections/200820298-Error-Pages. Accessed May 29, 2018.

[16] CLOUDFLARE DAMON. 1012 error: Access denied. Cloudflare support page: https://support.cloudflare.com/hc/en-us/articles/200171986-1012-Error-Access-Denied, Jan. 2017.

[17] CRANDALL, J. R., ZINN, D., BYRD, M., BARR, E., AND EAST, R. ConceptDoppler: A weather tracker for Internet censorship. In *Computer and Communications Security* (Alexandria, VA, USA, 2007), ACM, pp. 352–365.

[18] DALEK, J., HASELTON, B., NOMAN, H., SENFT, A., CRETE-NISHIHATA, M., GILL, P., AND DEIBERT, R. J. A method for identifying and confirming the use of URL filtering products for censorship. In *Internet Measurement Conference* (2013), ACM.

[19] ENSAFI, R., FIFIELD, D., WINTER, P., FEAMSTER, N., WEAVER, N., AND PAXSON, V. Examining how the Great Firewall discovers hidden circumvention servers. In *Internet Measurement Conference* (New York, NY, USA, 2015), ACM, pp. 445–458.

[20] EUROPEAN COMMISSION. General data protection regulation (GDPR). Regulation (EU) 2016/679, L119, May 2016.

[21] GEBHART, G., AUTHOR, A., AND KOHNO, T. Internet censorship in Thailand: User practices and potential threats. In *European Symposium on Security & Privacy* (2017), IEEE.

[22] GONLAG, M. What are the types of threats? Cloudflare Support page: https://support.cloudflare.com/hc/en-us/articles/204191238-What-are-the-types-of-Threats, May 2018. Accessed May 29, 2018.

[23] GROLL, E. How Washington helps Tehran control the Internet. *Foreign Policy* (Jan. 2018). http://foreignpolicy.com/2018/01/04/how-washington-helps-tehran-control-the-internet/.

[24] HANNAK, A., SOELLER, G., LAZER, D., MISLOVE, A., AND WILSON, C. Measuring price discrimination and steering on e-commerce web sites. In *Proceedings of the 2014 Conference on Internet Measurement Conference* (New York, NY, USA, 2014), ACM, pp. 305–318.

[25] HERN, A., AND WATERSON, J. Sites block users, shut down activities and flood inboxes as GDPR rules loom. *The Guardian* (May 2018). https://www.theguardian.com/technology/2018/may/24/sites-block-eu-users-before-gdpr-takes-effect.

[26] HILL, R. US websites block netizens in Europe: Why are they ghosting EU? it's not you, it's GDPR. *The Register* (May 2018). https://www.theregister.co.uk/2018/05/25/tronc_chicago_tribune_la_times_gdpr_lock_out_eu_users/.

[27] JOHNSON, D. L., PEJOVIC, V., BELDING, E. M., AND VAN STAM, G. Traffic characterization and internet usage in rural africa. In *Proceedings of the 20th International Conference Companion on World Wide Web* (New York, NY, USA, 2011), WWW '11, ACM, pp. 493–502.

[28] KARR, T. Net neutrality violations: A brief history. *Free Press* (Jan. 2018). https://www.freepress.net/our-response/expert-analysis/explainers/net-neutrality-violations-brief-history.

[29] KHATTAK, S., FIFIELD, D., AFROZ, S., JAVED, M., SUNDARESAN, S., PAXSON, V., MURDOCH, S. J., AND MCCOY, D. Do you see what I see? Differential treatment of anonymous users. In *Proceedings of the Network and Distributed System Security Symposium (NDSS)* (2016).

[30] KHATTAK, S., JAVED, M., ANDERSON, P. D., AND PAXSON, V. Towards illuminating a censorship monitor's model to facilitate evasion. In *The 3rd USENIX Workshop on Free and Open Communications on the Internet* (2013), USENIX.

[31] LOMAS, N. WTF is GDPR? *TechCrunch* (Jan. 2018). https://techcrunch.com/2018/01/20/wtf-is-gdpr/.

[32] LOWE, G., WINTERS, P., AND MARCUS, M. L. The great DNS wall of China. Tech. rep., New York University, 2007.

[33] MARCZAK, B., WEAVER, N., DALEK, J., ENSAFI, R., FIFIELD, D., MCKUNE, S., REY, A., SCOTT-RAILTON, J., DEIBERT, R., AND PAXSON, V. An analysis of China's "Great Cannon". In *Free and Open Communications on the Internet (FOCI)* (2015), USENIX.

[34] MAURER, T., AND MORGUS, R. Stop calling decentralization of the internet "balkanization". *Slate* (Feb. 2014). In the Future Tense blog: http://www.slate.com/blogs/future_tense/2014/02/19/stop_calling_decentralization_of_the_internet_balkanization.html.

[35] MIKIANS, J., GYARMATI, L., ERRAMILLI, V., AND LAOUTARIS, N. Detecting price and search discrimination on the internet. In *Proceedings of the 11th ACM Workshop on Hot Topics in Networks* (New York, NY, USA, 2012), ACM, pp. 79–84.

[36] MIKIANS, J., GYARMATI, L., ERRAMILLI, V., AND LAOUTARIS, N. Crowd-assisted search for price discrimination in e-commerce: First results. In *Proceedings of the Ninth ACM Conference on Emerging Networking Experiments and Technologies* (New York, NY, USA, 2013), ACM, pp. 1–6.

[37] NABI, Z. The anatomy of web censorship in Pakistan. In *Presented as part of the 3rd USENIX Workshop on Free and Open Communications on the Internet* (Berkeley, CA, 2013), USENIX.

[38] PARK, J. C., AND CRANDALL, J. R. Empirical study of a national-scale distributed intrusion detection system: Backbone-level filtering of HTML responses in China. In *Distributed Computing Systems* (2010), IEEE, pp. 315–326.

[39] PARRILLA, D. GDPR for lazy people: Block all European users with Cloudflare Workers. `https://apility.io/2018/05/25/gdpr-lazy-block-european-users-cloudflare-workers/`, May 2018.

[40] PAUKNER, P. Diese Kultur ist in Deutschland leider nicht verfgbar. *Süddeutsche Zeitung* (2013). `http://www.sueddeutsche.de/digital/streit-zwischen-youtube-und-gema-diese-kultur-ist-in-deutschland-leider-nicht-verfuegbar-1.1584813`.

[41] SFAKIANAKIS, A., ATHANASOPOULOS, E., AND IOANNIDIS, S. CensMon: A web censorship monitor. In *Free and Open Communications on the Internet* (2011), USENIX.

[42] SINGH, R., NITHYANAND, R., AFROZ, S., PEARCE, P., TSCHANTZ, M. C., GILL, P., AND PAXSON, V. Characterizing the nature and dynamics of Tor exit blocking. In *USENIX Security* (Aug. 2017).

[43] UNBLOCKVIDEOS.COM. YouTube region restriction statistics, check YouTube video restrictions online. `https://unblockvideos.com/youtube-video-restriction-checker/`. Accessed May 28, 2018.

[44] VERKAMP, J.-P., AND GUPTA, M. Inferring mechanics of web censorship around the world. In *Free and Open Communications on the Internet* (2012), USENIX.

[45] VISSERS, T., NIKIFORAKIS, N., BIELOVA, N., AND JOOSEN, W. Crying wolf? On the price discrimination of online airline tickets. In *7th Workshop on Hot Topics in Privacy Enhancing Technologies (HotPETs 2014)* (July 2014).

[46] WINTER, P., AND LINDSKOG, S. How the Great Firewall of China is blocking Tor. In *Free and Open Communications on the Internet* (Bellevue, WA, USA, 2012), USENIX.

[47] WIRED STAFF. Solution for slashdot effect? *Wired* (10 2004). `https://www.wired.com/2004/10/solution-for-slashdot-effect/`.

[48] XU, X., MAO, Z. M., AND HALDERMAN, J. A. Internet censorship in China: Where does the filtering occur? In *Passive and Active Measurement Conference* (Atlanta, GA, USA, 2011), Springer, pp. 133–142.

[49] ZHELEVA, M., SCHMITT, P., VIGIL, M., AND BELDING, E. The increased bandwidth fallacy: Performance and usage in rural zambia. In *Proceedings of the 4th Annual Symposium on Computing for Development* (New York, NY, USA, 2013), ACM DEV-4 '13, ACM, pp. 2:1–2:10.

[50] ZITTRAIN, J., AND EDELMAN, B. G. Internet filtering in China. *IEEE Internet Computing 7*, 2 (Mar. 2003), 70–77.