Digital Discrimination of Users in Sanctioned States: The Case of the Cuba Embargo

Anna Ablove, Shreyas Chandrashekaran, Hieu Le, Ram Sundara Raman, and Reethika Ramesh, *University of Michigan;* Harry Oppenheimer, *Georgia Institute of Technology;* Roya Ensafi, *University of Michigan*

https://www.usenix.org/conference/usenixsecurity24/presentation/ablove

This paper is included in the Proceedings of the 33rd USENIX Security Symposium.

August 14–16, 2024 • Philadelphia, PA, USA

978-1-939133-44-1

Open access to the Proceedings of the 33rd USENIX Security Symposium is sponsored by USENIX.

Digital Discrimination of Users in Sanctioned States: The Case of the Cuba Embargo

Anna Ablove* Shreyas Chandrashekaran* Hieu Le* Ram Sundara Raman*
Reethika Ramesh* Harry Oppenheimer† Roya Ensafi*

*University of Michigan †Georgia Institute of Technology

Abstract

We present one of the first in-depth and systematic end-user centered investigations into the effects of sanctions on geoblocking, specifically in the case of Cuba. We conduct network measurements on the Tranco Top 10K domains and complement our findings with a small-scale user study with a questionnaire. We identify 546 domains subject to geoblocking across all layers of the network stack, ranging from DNS failures to HTTP(S) response pages with a variety of status codes. Through this work, we discover a lack of user-facing transparency; we find 88% of geoblocked domains do not serve informative notice of why they are blocked. Further, we highlight a lack of measurement-level transparency, even among HTTP(S) blockpage responses. Notably, we identify 32 instances of blockpage responses served with 200 OK status codes, despite not returning the requested content. Finally, we note the inefficacy of current improvement strategies and make recommendations to both service providers and policymakers to reduce Internet fragmentation.

1 Introduction

Internet fragmentation is of increasing concern across the world, with several countries proposing locale-based content-restricting legislation in recent years [17, 22, 32]. The resulting rise of digital inequity places democracy under threat by restricting the free flow of information; still, there remains a shortage of data-driven studies that explore this problem. Existing work focuses on inaccessibility due to nation-state censorship [28, 45, 47, 57] or infrastructure limitations [11], but there is increasing anecdotal evidence of the detrimental effects of server-side geodiscrimination, or *geoblocking*. Reasons for this geoblocking remain understudied [2, 35, 49], especially with regards to the role of economic sanctions in causing digital discrimination.

Digital discrimination resulting from sanctions has seen a notable uptick and garnered significant global attention in recent years, following U.S. sanctions against regions such as Iran, Syria, North Korea, and Cuba [41]. For example, service providers such as Amazon Web Services and Slack recently removed support immediately after the U.S. government announced a novel sanctions program in Russia [15], leading to the Office of Foreign Assets Control (OFAC) [40] receiving immediate backlash. The agency had to issue a subsequent clarification to the sanctions to emphasize the government's commitment to "support the flow of information." This incident highlights the critical lack of clarity between policymakers and service providers, and mirrors similar disjoint responses in other U.S. sanction deployments.

In this paper, we investigate web content inaccessibility as a result of U.S. sanctions. As sanctions are often applied to unstable or repressed regions, engaging in this type of work has inherent challenges, including the ethical risks associated with interacting with in-region contacts and difficulties with obtaining a vantage point (VP) that accurately captures a residential perspective. Furthermore, even with an appropriate VP, much care is needed to attribute connection errors accurately. Prior work on server-side blocking highlights these complexities, noting compounding factors, including DDOS prevention, IP reputation, and distinguishing geoblocking from censorship [2, 30, 35, 49].

We present one of the first in-depth and systematic investigations into geoblocking in the context of embargo sanctions, prioritizing clarity to end-users. Specifically, we explore the following questions: What popular domains and content categories are being geoblocked? How is geoblocking implemented at the network layer, and what are the effects on measurement-level transparency? Correspondingly, how is geoblocking experienced on the ground, and does it promote user-facing transparency? These questions are crucial to identify the gaps in understanding between policymakers and web service providers and to fully characterize the digital effects of embargo sanctions.

We chose to perform this study on Cuba, primarily because we could ensure both safety and ethics compliance of the highest order. Since Cuba has a relatively higher civil-liberty ranking than other sanctioned states [23] and no active con-

flicts, the potential risks to participants of this work are minimized. Additionally, imposed in the 1960s, the U.S. embargo on Cuba is the longest running in history [16, 46]. As such, the totality of the development of Cuba's Internet ecosystem has been in the shadow of these regulations; still, there is an absence of clear characterization of the digital effects of the embargo. Having identified Cuba as our location of study, we gather knowledge from users on the ground in Cuba using a small-scale user study to gain insight into their perspective of the day-to-day impact of U.S. sanctions. We then use these results to inform our comprehensive network measurements of popular domains to identify geoblocking.

We test 10,093 domains, drawn from the Tranco Top 10K [31] domains and augmented with those highlighted in our small-scale user study in Sec. 4. Our measurements are informed by the valuable perspectives shared by our respondents and guided by the desire to rigorously ascertain how domains are geoblocked. We take extensive measures to identify geoblocking across different layers of the network stack, conducting TCP and TLS traceroute measurements to localize failure points and manually examining relevant HTTP(S) response pages. This gives us insight into both measurement and user-facing transparency of different geoblocking implementations. Throughout this process, we uphold the highest ethical standards to minimize the risk to our respondents and the independent collaborator who provided us with a residential VP, with further details in Sec. 3.

We find 546 domains within categories such as Technology, Business, and Economy & Finance, are implementing geoblocking, including popular online services such as Spotify, MailChimp, and CourseHero. Spotify in particular hosts several podcasts discussing the U.S. embargo on Cuba [25], [1, 3, 25], yet is not accessible for Cuban netizens, pointing them to the page in Fig. 1. Furthermore, we identify 17 diverse geoblocking implementations, ranging from DNS failures to HTTP(S) response pages with a variety of status codes. Alarmingly, geoblocking implementation is also diverse among top domains such as `tiktok.com` and `mathway.com` that are hosted by popular Content Delivery Networks (CDNs) (*e.g.* Amazon, Cloudflare, Akamai); 7.7% of geoblocking CDN hosted domains serve blockpages with a 200 OK. This further highlights the lack of measurement-level transparency for geoblocking. In terms of user-facing transparency, we find that 88% of geoblocked domains do not serve informative notice of why they are blocked.

By highlighting the nonuniform nature of geoblocking implementations and lack of transparency at both the measurement and user-facing levels, as well as directly integrating the perspective of everyday Cuban Internet users, our study demonstrates a disconnect between policymakers and service providers regarding the scope and intention of the U.S. embargo on Cuba. We believe that there is a pressing need for policymakers to clarify existing sanctions and assist service providers in understanding their legal responsibilities. These actions, taken in tandem with service providers' commitment to standardizing and clarifying geoblocking implementations, represent a step towards reducing digital discrimination. We hope to provide the insights necessary to begin catalyzing these changes and inspire data-driven future work.

Figure 1: **Spotify Blockpage in Cuba.** While Cuban netizens see an informative blockpage when visiting Spotify, to reach this page we had to manually follow a redirect link from a 301 redirect response. This is an example of the dichotomy between measurement and user-facing transparency.

2 Background & Related Work

Civil society and technical organizations have noted how sanctions could limit access to information in states including Cuba [5, 12]. In 2021, Time magazine reported that technology-based companies severely and abruptly restricted their products from use in Cuba due to the volatile nature of U.S. sanctions, leaving Cubans who relied on these products for their livelihoods to scramble and find alternatives [9]. UC Santa Barbara's Media Fields Journal provided additional accounts that corroborated the capricious and nontransparent product restrictions that resulted from the confusion about embargo policies [29]. There have been few data-driven studies that focused on geoblocking in Cuba. One such work, McDonald *et al.* [35], found 66 of the top 10K Alexa domains unreachable from Cuba.

History of Internet Access in Cuba. Despite being one of the first countries in the Caribbean region to gain access to the Internet in the late 1990s, Cuba's connectivity and infrastructure development never got off the ground due to the communist government's passage of laws aimed at limiting the free flow of information [4]. The early 2000s brought a ban on ownership and use of computers and cellphones that stymied even more growth as the rest of the world progressed into the Internet age [21]; its repeal in 2008 and the subsequent building of a sub-sea cable connecting Cuba to Venezuela in 2013 led to citizens being able to use the Internet [11]. To further regulate Internet access, Cuba merged its existing telecommunications providers to make the only legal state ISP, Empresa de Telecomunicaciones de Cuba S.A. (ETECSA) [20].

The popularity of the Internet grew throughout the 2010s, leading to ETECSA offering 3G/4G access plans by late 2019 [23, 34]. Prior work investigating network conditions in Cuba found that most outgoing traffic is served via the aforementioned sub-sea cable instead of the slower satellite chain alternative [11]. However, Internet connections still remain slow to this day [20]. Today, Internet access has permeated throughout the insular nation; in 2021, it was estimated that 72% of the Cuban population actively uses the Internet [23]. The Cuban Internet experience has been significantly shaped by government censorship and embargo restrictions, especially those placed on it by the U.S.

2.1 Sanctions

Purpose and Implementation. Sanctions are used by governments to apply pressure to target nations [7]. Policymakers design sanctions to achieve broad goals, such as wholesale regime change, or limited goals, such as specific policy changes. They hope that experiencing economic pain will lead citizens to lobby their governments to change course, elites to put pressure on their peers to adjust, or leaders to anticipate unrest and bend to external pressure [24, 26, 38]. Successful sanctions require cooperation between private firms and national governments.

U.S. Sanction Compliance. Businesses looking to provide products or services in embargoed or sanctioned countries must either apply for and receive specific authorization from the U.S. Department of Treasury or be a member of a list of allowed products and service categories promulgated by the Bureau of Industry and Security (BIS) [41]. These two bodies work together with the Office of Foreign Assets Control (OFAC) to ensure compliance with these regulations; any businesses found in violation can be fined up to 1M USD depending on the scale and severity of the transgressions [41]. Under OFAC's mandate, offering digital services of any kind in comprehensively sanctioned states requires either a specific or general license [51].

While still requiring OFAC approval (via a general license) based on the technologies used [33], free services do not have to undergo financial approval. Still, some technologically advanced software, such as those with encryption above a certain complexity, remains completely interdicted [41]. Covering less complex software, the U.S. State Department recommended the first general license for digital services "to allow Iranians to download free, mass-market software" in 2009 [56] and released statements supporting the free flow of information to this sanctioned region [53].

Noncompliance Risks. Sanctions, then, present a significant regulatory risk for businesses. Faced with a choice, they could conduct business as usual and pay continued compliance costs to remain within legal boundaries. Alternatively, firms can de-risk from markets even if it is permissible to operate there [19]. Regarding the negative impacts of sanctions, the U.S. Treasury Department recognizes that "de-risking hampers the unencumbered flow of development funding, as well as humanitarian and disaster relief" [55]. Even if sanctions do not explicitly forbid an action, they create compliance costs for doing business in the sanctioned country.

The U.S. Embargo on Cuba and Exemptions. The Cuban embargo was initially proposed by President Kennedy in 1962 and expanded by Presidents Johnson and Reagan, finally formalized with the Helms-Burton Act of 1996 [16, 46]. The goal of the embargo was to pressure Cuba to return to a democratic form of government by cutting off imports from and exports to the country [42]. In recent times, these regulations have seen significant change [8, 16, 42].

While a blanket sanction on Cuba exists, the U.S. government has attempted to clarify how sanctions apply to digital services along with their potential exemptions. Internet-based services are covered by a general license in US 31 CFR §515, first issued in 2010 addressing "the exchange of communications over the Internet." General exemptions exist for software services and fee-based Internet services (*e.g.* e-mail, VOIP), providing Internet connectivity, or travel to set up digital services [54]. The general license makes an additional carve out allowing services "widely available to the public at no cost to the user" to be provided to normally prohibited officials of the Cuban Government or Cuban Communist party, although it does not include the same language for entities on the State Department's Cuba Restricted List. Additionally, OFAC and BIS published a fact sheet in 2021 [40], detailing the government's commitment to protecting the fundamental freedoms of citizens of Cuba and outlining relevant categories of services potentially subject to exemption by reemphasizing some of the aforementioned excerpts of the CFR, as well as "software design, business consulting, and IT management and support (including cloud storage)," and "certain Internet-based courses" [54].

2.2 Geoblocking

Before proceeding, we find it prudent to discuss and disambiguate a few related terms mentioned in this work. *Censorship* typically refers to nation-state actors denying access to particular Internet services or groups of websites, while *server-side blocking* refers to service operators denying access to users based on some attribute of the user's connection. *Geoblocking* is a type of server-side blocking conducted based on a user's location. A *blockpage* is a page served instead of regular content, ranging from a blank page or nondescript error message to an informative explanation regarding the inaccessibility. While geoblocking can present with a blockpage, this is not always the case as certain implementations can cause errors before the page is served [49].

Existing Geoblocking Studies. As the fragmentation of the

Internet has gained in prominence, geoblocking has emerged as a recent topic of study. Tschantz *et al.* [52] analyzed geoblocking by querying Cloudflare-hosted domains in the now deprecated Alexa Top 1M websites using virtual private servers (VPSes) in several countries, identifying geoblocked domains as those whose contain standardized indicators of geoblocking within CDNs. They found 524 out of 77K domains served such a blockpage.

McDonald *et al.* [35] conducted a wide-scale set of measurements on geoblocking, querying the Alexa Top 10K websites from thousands of VPSes in 177 countries. They focused on geoblocking implemented by Content Delivery Networks (CDNs), finding Cuba experienced geoblocking on 66 out of 10K and 165 out of the Alexa Top 1M websites, lagging behind Syria, Iran, and Sudan.

Afroz *et al.* [2] leveraged similar techniques and conducted measurements to study server-side blocking in Africa, Pakistan, and Ukraine. In an attempt to evaluate a more expansive range of geoblocking throughout the network stack, they used heuristics surrounding the distance traveled by an outgoing packet along the network path to localize the point of failure and determine its proximity to the server.

Kumar *et al.* [30] investigated geoblocking implemented via the Google Play Store, developing a semi-automated tool to download apps as seen by users in 26 countries. They found 3,672 out of 5,684 apps (64.6%) were geoblocked in at least one location by service providers, a departure from previous conclusions of sparse and category-specific geoblocking by McDonald *et al.* and Afroz *et al.*

Decoupling Geoblocking from Censorship. Previous work relied on the presence of blockpages or numerical heuristics to identify geoblocking [2, 35, 52]. In a bid to remedy these shortcomings, Ramesh and Sundara Raman *et al.* [49] studied both censorship and geoblocking during the Russian invasion of Ukraine with an eye towards an accurate, conservative decoupling mechanism. They developed and open-sourced a measurement tool [13] that checks the accessibility of a domain across different network layers. Additionally, they run traceroute measurements to differentiate between geoblocking and censorship, capitalizing on the disparate locations of connection failure in each case. We leverage and extend these methods, as discussed in Sec. 5.

Our Work in Perspective. We present an in-depth and systematic investigation into geoblocking through the lens of embargo sanctions, as well as conduct the first small-scale user study capturing the impact of these sanctions on everyday Internet users. We note the limited prior work on Cuba focuses on characterizing general service availability [11], and the only other geoblocking study including Cuba from McDonald *et al.* [35] limits its scope to CDN blockpages. In comparison, we consider geoblocking implementations across network layers, resulting in outcomes like failures in DNS resolution to blockpages served with 200 OK responses. From a technical network measurement perspective, we not only utilize state-of-the-art methodology [13, 49], but also necessarily extend it to deal with extremely poor network conditions in Cuba by adding a configurable sleep time between measurements. We have open-sourced these changes [13].

3 Ethics

We consulted with our Institutional Review Board (IRB) and received IRB exemptions for both our small-scale user study (Sec. 4) and network measurement study (Sec. 5). We further consider the design of our studies and their potential risks by reaching out to the general counsel, colleagues at our institutions, and collaborators with expertise in researching censorship. Furthermore, we aim to uphold the ethical principles in the Menlo Report [6], which is inline with prior work [27, 39, 44, 48, 59]. These principles impact the design and deployment of our studies.

The first principle is *respect for persons*. We carefully engage in secure, small-scale outreach through established contacts, with our resulting 10 respondents skewing more technically literate. This ensures that they can give informed consent and understand the potential risks involved. Before proceeding through the questionnaire, we provided a detailed disclosure form, noting participation was entirely optional.

The second principle is *beneficence*. We protected our respondents' anonymity by not collecting any personally identifiable information (PII), such as names or email addresses. For our network measurements, completely disjoint from the questionnaire, an Internet freedom community member with years of experience related to network traffic analysis of Cuba explicitly consented to provide the Cuban VP to facilitate the study. Due to an inability to identify another similarly experienced contact, we limited ourselves to the use of the single VP. Using this VP, we conducted measurements to determine the accessibility of domains within Cuba. A network observer could attribute regular requests to censored domains as being from the VP owner. We balance these risks with the understanding that our work could directly aid Cuban citizens in uncovering the extent of the digital effects of the embargo.

The third principle is *justice*. In Sec. 6, our findings highlight the inefficacy of current improvement strategies to reduce the negative impact of the embargo on the free flow of information in Cuba. As our work pushes for better standardization of geoblocking implementation to minimize consequences, this directly benefits all of our participants. The fourth principle is *respect for law and public interest*. To adhere to the embargo on Cuba, we conducted the small-scale user study with volunteers only (*i.e.* without financial compensation).

4 Small-Scale User Study

In this section, we explain how we conduct our small-scale user study to characterize the impact of the U.S. embargo sanctions on Cubans from a first-hand user perspective, as well as identify exact websites and services that are blocked for our respondents. Extracting these experiences and struggles helps inform our network measurements in Sec. 5 and our analysis of geoblocking in Sec. 6. The full questionnaire is provided in Appendix A.1.

4.1 Motivation and Methods

Despite ample news coverage surrounding the downstream effects of U.S. embargo sanctions on Cuban Internet, such as those detailing that Cubans cannot access major services such as MailChimp and OpenSea [9], there is a lack of data-driven studies that capture the embargo's technical impact from a user perspective. We seek to explore how the embargo affects users on the ground.

There are innate challenges associated with carrying out our questionnaire. First, we had to ensure through our contacts that our questionnaire platform, TypeForm, was accessible to Cuban respondents. Second, we had to solicit participation by conducting a small-scale ethical outreach through established contacts to mitigate respondent risks. As such, we leveraged personal and NGO contacts to distribute our questionnaire to 10 Cuban Internet users. We note that our selected respondents skew toward being technologically literate. Since our study centers around user perspectives of sanction-related geoblocking and its technical outcomes and impact, in addition to the ethical benefits detailed in Sec. 3, this skew is advantageous because technical respondents are more likely to provide relevant and accurate details. Additionally, we note that the limited and guarded mode of outreach ensures all users taking our questionnaire are fully aware of the risks associated with participation.

We begin the questionnaire with a statement detailing both respondent risk and the measures we take to ensure anonymity and privacy. Our first section collects anonymous general demographic and Internet usage information; we next ask about notable blocked web services and categories and their specific blocking indicators, such as blockpage screenshots. Finally, we ask respondents to rate how the inaccessibility of various categories of web services affects their daily lives to capture the human impact of geoblocking. Responses were collected in English (4) and Spanish (6), the latter translated to English with Google Translate and DeepL [58]. We cross-validate and perform manual lookups to verify correctness.

We briefly characterize user responses using the following approaches. For polar questions, as well as those answered with the Likert scale, we explored aggregated metrics. For long-form questions, in which nine out of 10 respondents provided answers, we apply a free-form coding approach and find responses largely fall under two broad themes: the impact of sanctions on respondents' daily lives and their perspectives on sanctions policy; we highlight excerpts to characterize the depth of our respondents' perspectives.

4.2 Findings

All respondents reside in Cuba and the majority are between the ages of 26–35. Their familiarity with Internet technologies varies from daily Internet users to software professionals; five respondents identified as software developers or network administrators, one as a government official, and the remaining noted various other occupations.

Geoblocking Characterization. Respondents identified 36 domains as subject to geoblocking and pointed us to a community-maintained repository with 95 domains purportedly geoblocking Cuban users. We augment our test list with these domains in Sec. 5, with results in Sec. 6. Regarding common user-facing instances of geoblocking, all 10 respondents report encountering 403 Forbidden pages, and nine identify location-based language such as "Not available for users in your region." R2 additionally highlights non-standard blockpages, adding in an optional response,

> "The page loads, but not all the content. Vital bits of it say it's not available from my location." (R2)

Only five respondents attribute ambiguous browser error feedback (*e.g.* `No Internet`) to server-side blocking. Among the categories of services subject to geoblocking, Communication Tools (8 respondents), File-sharing (6 respondents), and E-commerce (6 respondents) were the most common.

Notably, although we sought instances of geoblocking rather than censorship, respondents still attributed blocking in categories like Political Criticism and Human Rights Issues to geoblocking when they are more likely to be censored. In fact, one respondent identified `cibercuba.com` as geoblocked, but we confirm that it is censored in Sec. 6. This suggests a lack of clarity surrounding the disambiguation of geoblocking and local censorship from an end-user perspective and a generalized lack of user-facing transparency.

Impact on Daily Life. Fig. 2 displays aspects of our respondents' daily lives impacted by sanctions. While some respondents highlight services with financial transactions that more directly fall under the scope of sanctions, several also note propagating impacts on their careers. For instance, R7 and R8 discuss barriers to information and the isolated nature of a Cuban developer work:

> "I do not have access to many tools that are blocked in my country, affecting my profession, limiting my knowledge." (R7) "...it makes it impossible for us to create projects that can be used internationally." (R8)

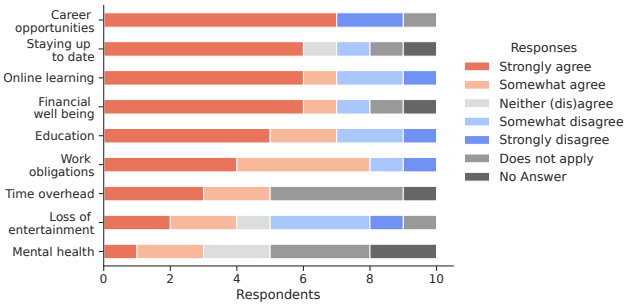

Figure 2: **Impact on Daily Life.** Respondents reveal how embargo sanctions impact their daily lives, with "Career opportunities," "Staying up to date," and "Online learning," noted as being the most negatively impacted.

Similarly, R9 complains at being excluded from the global community, *"The blocks to services that in the world are practically 'natural,' in my case they are impossible to use."*

Further, this isolation impacts respondents beyond their livelihoods. R4 notes *"The impact on my life is daily, whether it is to download and use apps, to pay, for work meetings and many more."* More broadly, R1, a tourist guide, notes, *"It is another form of segregation."*

In terms of the availability of substitutes for sanctioned services, seven out of 10 respondents were unaware of any local alternatives. For those aware of alternatives, the gap remained apparent.

> "There are some that we have had to create ourselves, although their capabilities are not as good as the originals." (R8)

Thus, it is unsurprising that nine of our respondents resort to circumvention tools to access sanctioned services. However, we emphasize the tech-savvy skew of our respondents and note that this is not generalizable to the greater Cuban populace. Notably, there are downsides to using these tools, such as safety concerns, as well as time and financial costs,

> "I expose myself to being discovered as it is considered illegal in my country." (R7) "[It] slows down my day to day, and makes my connection more expensive." (R2)

User Perspectives on Geoblocking. Five respondents offer perspectives on sanctions directly and indicate perceived unintended targets of their negative impacts. R7 fully disregards any harm to the government from sanctions, *"Blocked sites only affect the people who need the knowledge, not the government;"* while R10 takes a more moderate approach, criticizing the lack of precision of sanction effects,

> "I can understand the goal of restriction regarding to governments. But this should be better targeted to not affect people not directly related to the sanctioned government or state." (R10)

Similarly, R8 details their perspective on the embargo harming development,

> "In theory, these sanctions should only affect the government, but they affect all citizens directly, greatly limiting our chances of prospering and being able to develop like any other citizen in the world." (R8)

Takeaways. We observe consistent levels of frustration and claims about the negative impact of the U.S. embargo on Cubans. Our respondents are especially cognizant of how sanctions have affected their professional livelihoods, lamenting their loss of opportunities for international jobs and financial growth. Our small-scale user study informs our systematic measurement approach by augmenting our test list in Sec. 5. Notably, for Sec. 6, it underscores the importance of studying user-facing issues, like geoblocking transparency.

5 Measurements

Our goal is to identify and study instances of geoblocking in Cuba. However, due to the sensitive nature of U.S.-Cuban relations and poor network conditions inside Cuba, we apply nuanced and rigorous methods, mitigating potential data loss while adhering to ethical principles as outlined in Sec. 3.

Fig. 3 provides an overview of our approach to measuring and identifying geoblocking. Our test list comprises of the Tranco Top 10K domains [31], which we augment with 36 domains mentioned by respondents in our questionnaire in Sec. 4, as well as 95 domains from the aforementioned community-maintained repository [18]. For our control measurements in Sec. 5.1, we rent three VPSes in data centers, two in the U.S. and one in the U.K. For our Cuba measurements in Sec. 5.2, we rely on a collaborator in Cuba and obtain a VP located in the state-owned and only public ISP Empresa de Telecomunicaciones de Cuba (ETECSA). Notably, prior research has shown that measurements using data center VPSes are not representative of true user experiences in networks experiencing interference [57]. Thus, we run our measurements from a residential machine. Additionally, we retest domains in Cuba that fail all DNS requests or fail with a server error associated with timing out, as these could be due to poor network conditions, as further detailed below.

Measurement Runs and VP Instability. Our measurements were conducted over the period of May 11–22 of 2023, with corresponding control and Cuba measurements conducted within 24 hours of each other to ensure consistency. Throughout the data collection period, the instability of the VP's network remained a primary point of concern. Due to both a lack of access and the challenging nature of obtaining additional

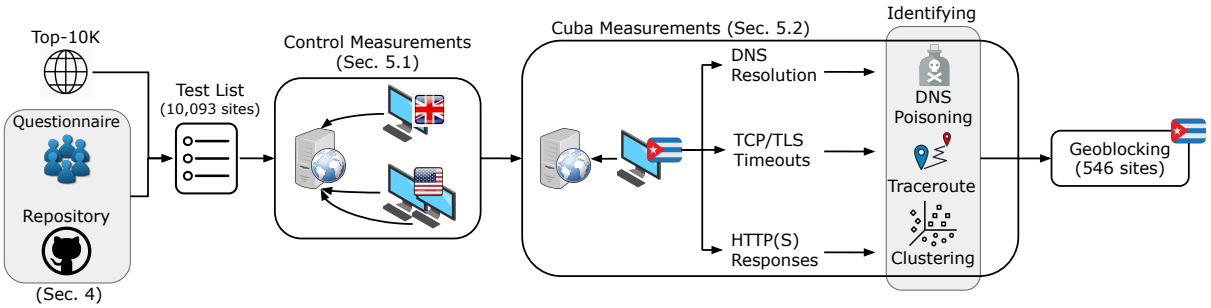

Figure 3: **Measurement and Geoblocking Identification Overview:** We first construct our test list with the Tranco Top 10K domains and augment it with domains from our small-scale user study and community-maintained repository. Then, we conduct control measurements from our three control VPSes. Next, we run Cuba measurements from our residential VP and identify instances of geoblocking across different layers of the network stack.

VPs, we operated under the necessary premise that loss of machine access could occur at any time.

Our initial measurements run 100 domains in parallel, with successive request batches spaced by one second. Due to the unreliability of the residential VP, in cases where test measurements fail all DNS resolution attempts or experience a server-level time out error, we retry each domain an additional time. We configure retest measurements with a lowered parallelization rate of 20 domains and increased batch spacing of 30 seconds. This allows us to distinguish slow network conditions from geoblocking. We update our open-source measurement software with this configurable wait time [13].

5.1 Control Measurements

Through our control measurements, we identify and remove inaccessible domains from our test list. This prepares the data as a benchmark for further analysis.

Our measurements begin in the DNS layer, where we perform three types of measurement for each domain. First, we query via our custom recursive DNS trace method, which enables us to function as the client by iteratively requesting each segment of the domain (*e.g.* for `a.example.com`, we query `.com`, then `example.com`, then `a.example.com`) until we reach the desired A-record. Next, we perform two rounds of public DNS resolution, via Cloudflare DNS (`1.1.1.1`) and Google DNS (`8.8.8.8`), the latter being the default DNS resolution server of the measurement machine. Cloudflare's nearest point of presence to Cuba is in the United States (Miami, Florida), and Cloudflare does not forward the Client IP information by default. Similarly, Google DNS connects to a server in the United States (Charleston, South Carolina).

We then attempt to establish a TCP connection with the web server using the IP address obtained during the DNS resolution. To clearly separate unavailability-related indicators from transient network failures, we retry failed TCP handshakes up to three times, within a given run of a domain. Once the TCP connection is established, we attempt to perform a TLS handshake, sending a TLS Client Hello with the Server Name Indication (SNI) field set to the domain, replicating browser behavior. After the TLS connection is established we send an HTTP GET request with the Hostname set to the test domain. We follow up to three redirects and store the response from the web server.

Removing Unreachable Domains. Only one domain's accessibility differed between the U.S. and U.K. machines, and we removed this domain from consideration. The majority of domains that fail to resolve consist of internal DNS domains used for name resolution within private networks, like `awsdns-37.org`, and internal CDN hosting domains not intended for direct access, like `cdnbuild.net` and `rackcdn.com`. In total, of our initial measurement test list of 10,093 domains, we found 8,537 passed control DNS runs and 6,396 passed control TCP. We use this sanitized list of 6,396 domains as a baseline for comparison in Sec. 5.2. We additionally provide a breakdown of our test list by category in Fig. 12 in the Appendix.

While we proceed with the Tranco list as it is a state-of-the-art domain ranking list robust against manipulation (*e.g.* fake traffic), we also takes steps to account for the domains lost in the sanitization process. To explain the 36% difference between the initial and final lists, we cite a known issue with the Tranco list. It is a composite of four existing top lists, and its constructors acknowledge that 10-50% of domains in these lists did not issue a 200 OK response code or were otherwise unreachable [31]. We further verify that the difference in our lists is not unexpected by performing a t-test for the difference in means between our sanitized test list and bootstrapped samples of the Tranco list, repeatedly sampling from a theoretical distribution parameterized by the four compositional components of the list. We find the difference is not statistically significant at a significance level of $\alpha = 0.05$, suggesting that our sanitized test list falls within the expected distribution of accessible and inaccessible domains.

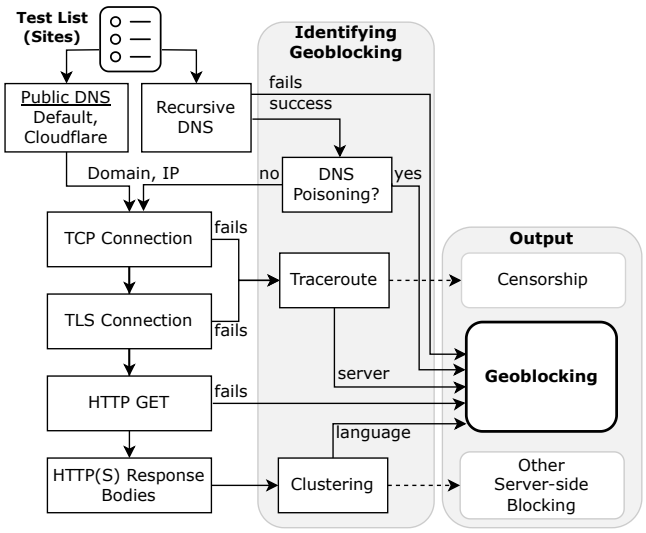

Figure 4: **Geoblocking Identification Pipeline:** We utilize various techniques to identify geoblocking and decouple it from censorship and other sever-side blocking at each layer of the network stack.

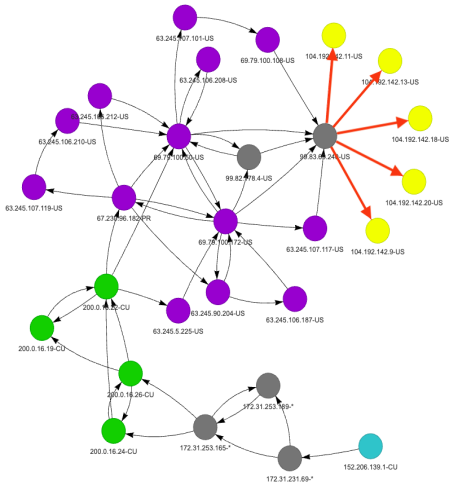

Figure 5: **TCP Traceroute Example.** Through our TCP traceroute measurements, we find Atlassian owned domains, `atlassian.com`, `atlassian.net`, `trello.com`, `atl-paas.net`, and `opsgenie.com`, failing upon reaching Amazon-AES (AS 14618). The red → indicates the hop where the SYN packet was dropped and nodes are colored by AS.

5.2 Identifying Geoblocking in Cuba

Fig. 4 illustrates our approach to measuring and identifying geoblocking in Cuba, broken down by layer-wise attribution. Throughout the following subsections, we detail how we decouple geoblocking from censorship.

Determining Layer of Failure. We conduct retests to compensate for poor network conditions in Cuba. Within the DNS and server-level (TCP, TLS, HTTP) failures, which never return an HTTP(S) response, we find nine cases where our retests have different outcomes from the initial measurement. We do the following to reconcile these differences: in the case of failures in different layers, we attribute the lower layer failure to transient error or path differences due to router load balancing and report the higher layer failure; in the case of failures in the same layer, we follow precedence orderings based on the informativeness of different errors.

5.2.1 DNS Geoblocking

We begin our DNS layer analysis by examining the output of our custom recursive DNS trace, as shown in Figure 4. Note, we group Google and Cloudflare together as public resolvers. Though we analyze all DNS results for DNS poisoning, we only consider outcomes from the recursive trace as potentially indicative of DNS geoblocking.

Analyzing for DNS Poisoning. To ensure the IP addresses we receive are correct and errors are localized to the precise location of the network stack, we examine them for DNS poisoning. We utilized MaxMind, GeoIP2 Tables, and Censys [14] to compare the autonomous system (AS) numbers and organizations between the 102 domains whose sets of returned IP addresses map to different ASes in the control and Cuba runs. In the public DNS resolution, we identify one suspicious case that returns a China Telecom IP address instead of an Amazon one via Google DNS resolvers from Cuba, suggesting blocking via the eDNS client subnet. The remaining successfully resolved IP addresses, along with their domains, move to the next step in our pipeline.

5.2.2 Web Server Geoblocking: Errors

Geoblocking can occur during the establishment of the TCP connection, TLS connection, or HTTP GET request. HTTP GET errors are easily identifiable as geoblocking; since we have already established the TLS session with the server, we can attribute errors such as HTTP truncated responses to our target web server for a given domain. In contrast, TCP and TLS errors necessitate a complex identification process. For TCP timeout outcomes, the TCP SYN packet could be dropped silently within Cuba, indicating IP-based censorship [49]; similarly, failures in the TLS layer could be caused by SNI-based censorship. Thus, we use traceroute measurements to localize the point of failure and decouple instances of censorship from geoblocking.

TCP Traceroutes. We conduct TCP traceroute measurements for 131 unique IP addresses corresponding to 95 domains that time out while establishing TCP connections, with an example shown in Fig. 5. In exactly one case, we find a termination very close to Cuba, indicating an instance of censorship: `eldiario.es` fails in AS 11960, the Internet exchange point

(IXP) through which all of Cuba's traffic propagates. Additionally, we successfully complete the TCP handshake with the server for seven of these 131 traceroutes. This can be caused by a variety of reasons, including servers handling the traceroute packets with short TTL values differently [37]; as well as routing differences due to the traceroute measurement necessarily being conducted with a different port number than the initial measurement, which has been shown to impact domain accessibility [10].

TLS Traceroutes. We also run traceroute measurements for 69 IP addresses corresponding to the 35 domains failing in the TLS layer. We find 35 traceroutes corresponding to 13 domains that terminate with either FIN or RST packet within Cuba's IXP (AS 11960), including `cibercuba.com` and `doubleclick.net`, which indicates censorship as opposed to server-side blocking.

Censorship Validation. In total, we find 14 censored domains. We corroborate our findings of censorship for seven of the domains discussed above with the OONI Explorer dashboard [43], which hosts in-country measurement data fine-tuned to capture Internet censorship. The remaining seven domains are not on the OONI test list.

Network Stages (5)	Measurement Outcomes (17)	# Domains
DNS Fails 37/546 **(6.8%)**	Failed iteration	33
	Failed AuthNS connect	2
	Manipulated IP	2
TCP Fails 97/546 **(17.7%)**	Timeout	95
	Network unreachable	1
	No route to host	1
TLS Fails 23/546 **(4.2%)**	Timeout	11
	Reset	9
	Fail	3
HTTP Fails 24/546 **(4.4%)**	Timeout*	6
	Reset*	3
	Truncated Response	6
	DNS fail in redirect*	10
HTTP Responses 395/546 **(72.3%)**	403 Forbidden	347
	451 Unavailable for Legal Reasons	9
	200 OK	32
	Others *e.g.* 404, 503	7
Geoblocked Domains		**546**

Table 1: **Geoblocked Domains.** We find explicit geoblocking signals for 546 domains. Percent failures are reported by stage on a per-domain basis. 30 domains block at multiple network stages; one domain has varying outcomes within the HTTP stage, which we denote with asterisks. We provide descriptions of HTTP response status codes in Table 3.

5.2.3 Web Server Geoblocking: Response Bodies

We now focus on HTTP(S) response bodies that we collected. Our goal is to identify geoblocking blockpages. However, there is much variety among these, especially between pages served by different CDNs [35] and domain owners. Blockpages served by CDNs are typically more uniform and standardized, whereas personalized blockpages served by individual domain owners vary widely in status codes and verbiage. Thus, we identify them through an iterative clustering process.

We leverage prior insight from Jones *et al.* [28] and McDonald *et al.* [35] and identify geoblocking patterns by comparing the length of Cuba measurement response bodies to their corresponding control response bodies. The intuition is that the geoblocking blockpages will have drastically different content. We first gather successful 200 OK responses from the control measurements and their corresponding responses in the Cuba measurements and use BeautifulSoup for text extraction. Then, we calculate a percentage change in response length for each extracted text pair. We identify potential blockpages as those whose lengths change by more than a certain threshold, which prior works empirically estimate at 30% [28, 49]. Next, since 403 and 451 response codes are known to be indicative of geoblocking [35, 49, 52], we also consider them as potential blockpages. In Sec. 6, we consider 403 and 451 response codes as *standardized* geoblocking implementations.

Using a threshold of 30% yields 932 domains with 1,062 unique HTTP(S) responses, in addition to 6 domains with 403 and 451 status codes below this threshold. We normalize the extracted text to remove interchangeable response qualities such as error ID numbers and url-strings and perform agglomerative clustering to group all responses with 90% unigram similarity. This results in 316 clusters, which we refer to as *fingerprints*. Examples of fingerprints are shown in Fig. 6. We classify fingerprints corresponding to responses with a 403 or 451 status code as geoblocking and examine the remaining fingerprints for explicit language-based indications of geoblocking. We detail the outcomes in Sec. 5.3.

Threshold Verification. To thoroughly identify responses containing indications of geoblocking, we test five linearly decreasing thresholds from 30% content difference down to 10% and carry out the above steps for each threshold. We find that using a threshold of 10% resulted in 97 additional false positive pages and only one more true positive. Thus, we choose to report results on our optimal value of 30% and include the singular additional true positive in our analysis for completeness; we further report the CDF of true and false positive HTTP(S) responses identified with respect to response length difference in Fig. 11 in the Appendix.

5.3 Data Characterization

Table 1 summarizes the diverse geoblocking implementations we identified in our Cuba measurements.

DNS Geoblocking. *We find 37 domains geoblocking in the DNS layer.* For two domains, we find their authoritative name servers engaging in DNS poisoning and returning `localhost`. Another two domains (*e.g.* `gitee.com` and `rt.ru`) reach their respective authoritative name servers but fail to connect.

For the remaining 33 of these domains, we fail to re-

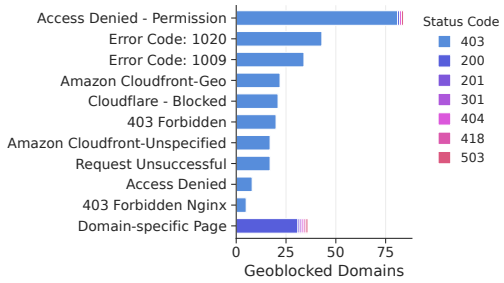

Figure 6: **Top 10 Fingerprints and Domain-specific Pages.** We show the top 10 fingerprints we find through our iterative clustering process and associated status codes served by domains. We bucket the personalized, domain-specific pages identified via language indicative of geoblocking.

trieve an IP address from the authoritative name server. To further investigate, we use our recursive DNS trace to locate the specific name server that failed to respond to our DNS query. Of these located name servers, we utilize MaxMind, GeoIP2 Tables, and Censys to confirm the ASes of 30 out of the 33 are directly associated with their corresponding domains. For instance, we find the name server `192.102.198.240` Intel-SC-AS (AS 4983) preventing access for three Intel-related domains (`01.org`, `intel.cn`, and `intel.com`). For the three domains where we were unable to definitely link the AS to the domain, we find two equivalent domains (`wdc.com` and `westerndigital.com`) and a government domain (`virginia.gov`). The equivalent domains both fail at the same server in Amazon-02 (AS 16509), and the government domain fails in a Verizon Business AS designated for customer usage and registered in Virginia.

Web Server Geoblocking Errors. *We find 97 domains implementing geoblocking via TCP handshake failures, and 23 domains via TLS connection failures. 24 domains were geoblocked at the HTTP GET stage.* During TCP connections, the most common failure we observe is connections timing out due to web servers not responding to TCP SYN packets. Analyzing traceroutes in the TCP layer, we find that 87 domains end within the same country as the destination IP address, 25 of which end inside the same autonomous system. Upon clustering on the AS of the last successful hop at the domain level, we find 10 domains reach Prolexic-DDOS-Network (AS 10794), 10 reach Level3 (AS 3356), eight reach Zayo-6461 (AS 6461), and seven reach Amazon-AES (AS 14618). Fig. 5 shows domains owned by Atlassian failing upon reaching Amazon-AES. In addition to blocking by organizational AS such as `bankofamerica.com`, which terminated in BankAmerica (AS 10794), we see the presence of transit networks such as Level3 (AS 3365), which blocks domains such as `adobe.com`, `barracuda.com`, and `citi.com`. We emphasize that the errors in these layers do not provide any transparency to the user about geoblocking.

Web Server Geoblocking Responses. *We identify 395 domains implementing geoblocking via blockpage HTTP(S) responses. Of these, 32 domains serve blockpages with 200 OK status codes.* Through our use of the iterative clustering and language-based methods to identify fingerprints, as described in Sec. 5.2.3, we find that while 356 domains serve status codes 403 and 451, the remaining 39 serve other various response codes including 200 OK, 404 Not Found, and 503 Service Unavailable. Fig. 6 depicts the 10 most common fingerprints associated with these domains, as well as an aggregation of domain-specific fingerprints.

We now focus our analysis on our geoblocking fingerprints. Recall we identified 316 unique fingerprints, of which we classified 100 as indicative of geoblocking. Of these, 71 fingerprints are classified as geoblocking because they contain responses served with 403 or 451 status codes. Notably, we find three domains matching these fingerprints, but returning non-403 or 451 response codes, which we also attribute as geoblocking. The other 29 fingerprints are classified as geoblocking because they contain indicative language (*e.g.* "Unavailable in your region"). We find these are all personalized to a given domain, except for domains `birmingham.ac.uk` and `sd.gov`, which serve the same fingerprint but with status codes of 503 and 200, respectively. We discuss geoblocking transparency afforded to users associated with these blockpages in Sec. 6 with examples in Table 2.

5.4 Manual Examination of Services

To better understand and characterize geoblocking, we manually examine geoblocked domains and their associated services to collect information that complements our findings.

Domain Exploration. It is common practice for companies to engage in embargo compliance disclosures on their websites. In order to identify these disclosing domains, two researchers independently visited all the 546 geoblocked domains from the U.S. on a university network and searched for privacy policies, terms and conditions, legal notices, and other compliance pages. Searches were conducted via a combination of browser and website-level keyword searches, as well as manual navigation. Relevant passages were extracted using keyword searches and manual readings. A list of keywords was maintained throughout, terms were iteratively added upon discovery, and domains were subsequently reexamined to ensure completeness. Some example phrases include the keywords "OFAC," "compliance," and "Cuba." All findings were cross-validated between researchers. We stress that this process was entirely manual, as prior work has shown the ineffectiveness of machine learning and automated approaches [36].

Free-tier Service Offerings. To characterize the blocking of types of services promoted for exemptions by the U.S. government (as noted in Sec. 2 and further discussed in Sec. 6), we map promoted types of services to relevant Cloudflare catego-

rizations. We identify the following categories: Technology, Information Technology, Business, Content Servers, Communication Tools, and Education. For the 388 geoblocked domains in these categories, we additionally collect information about free-tier service offerings during the manual domain exploration, as described above.

Country of Registration. We examine the country of registration for geoblocking services. We find 58 non-U.S. registered domains, all of which passed control measurements from the U.S. and U.K. While we cannot as clearly infer the reason for geoblocking for the non-U.S.-registered companies, six disclosed U.S. embargo compliance or Cuban inaccessibility, including `bmo.com` (Bank of Montreal) and `hootsuite.com` (Hootsuite), despite it being U.S. legislation. We discuss possible causes of non-sanctions geoblocking in Sec. 7.

Impact of Disclosures. Disclosures of sanctions compliance can further support our geoblocking identification methodology in Sec. 5.2, but have some restrictions. While a disclosure can help confirm geoblocking due to sanctions, the absence of one does not necessarily imply non-sanctions geoblocking. More critically, we want to acknowledge that disclosures do not help users in embargoed countries, as the websites themselves are inaccessible. This is particularly notable for domains like `webex.com`, which discusses the geo-availability of Webex services on its site, but provides no transparency to Cuban netizens, as it blocks via both the DNS and TCP layers. Further, even non-U.S. registered companies may comply with embargo sanctions, highlighting another component of the complexity of auditing geoblocking due to sanctions compliance. We further evaluate these effects in Sec. 6.

6 Results

In this section, we evaluate the 546 geoblocked domains to analyze the scope of blocking and compliance disclosures (Sec. 6.1), as well as user and measurement-facing transparency (Sec. 6.2), standardization (Sec. 6.3), and the inefficacy of current improvement strategies to minimize harm from U.S. sanctions (Sec. 6.4).

6.1 Geoblocking of Popular Domains

By Popularity. *We find that geoblocking is most prevalent across the Tranco Top 2K domains (27% of 546); we further observe that domains are geoblocked by diverse approaches across all ranks.* Fig. 7 shows a histogram of geoblocking of domains across the Tranco 10K+ domains. We observe responses served with 403 status codes like `att.com` (#337) and `ebay.com` (#97), and TCP timeouts, like `exacttarget.com` (#577) and `trello.com` (#299), are the most prevalent types of geoblocking approach throughout.

We verify that 56 of the 152 domains in the Tranco Top 2K (37.5%) provide some notice of embargo compliance on their website, compared to 111 of the 394 domains beyond the Top 2K (28.5%). Furthermore, we find that 38.6% of disclosing domains in the Top 2K mention Cuba specifically, compared to just 7.4% of domains beyond the Top 2K. This suggests that popular domains engage in more specific disclosures.

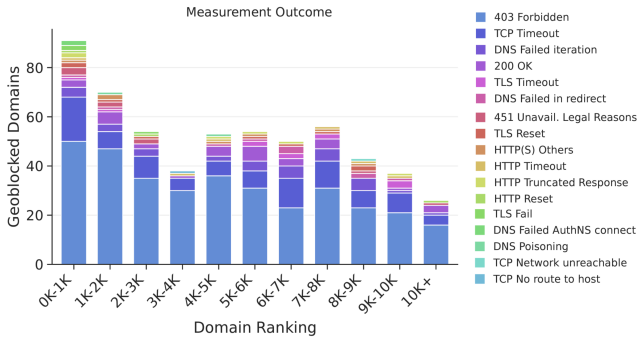

Figure 7: **Geoblocking vs. Popularity.** We show the number of domains geoblocked across our test list by Tranco rank and highlight outcomes from Table 1. Note, some test domains from our augmented test list fall outside the Tranco Top 10K.

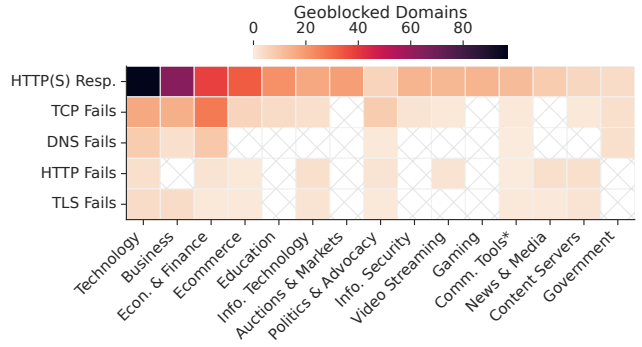

Figure 8: **Network Stages vs. Categories.** We show geoblocking by network stages for the Top 15 categories; and Communication Tools, an aggregation of Internet Phone & VOIP, Instant Messengers, Chat, and Forums, to compare with Sec. 4.

By Categories. *Technology, Business, and Economy & Finance are among the most geoblocked and often by 403 status codes, TCP timeouts, and DNS failures.* We use Cloudflare's Radar API to categorize each geoblocking domain; Fig. 8 details the presence of geoblocking across our Top 15 categories. We find that eight of these engage in various geoblocking methods over at least four network stages.

Technology domains, like `zoom.com`, implement geoblocking by returning a 403 status code more frequently than other methods, followed by Business and Economy & Finance related domains, such as `amplitude.com` and `geico.com`. In addition, these categories also engage in the highest amounts of geoblocking via TCP timeout.

Beyond the Top 15 categories (shown in Fig. 8), we find that

Transparency	Example Responses	Fingerprints	Domains	Disclosed
Informative	"This server denied by region", "Not available in your area", "IP address associated with a country or region currently subject to U.S. economic and trade sanctions"	48	79	59.5%
Vague	"Access Denied", "403 Forbidden", "Request blocked"	24	182*	30.2%
None	"There was a problem with the request", "Error"	28	135*	30.4%

Table 2: **Transparency of Geoblocking Fingerprints** We label each response as either "Informative," "Vague," or "None" to characterize user-facing transparency for blockpages. We further analyze these responses based on the number of fingerprints, geoblocked domains, and the proportion of disclosing domains. One domain served two different fingerprints (each in a different AS), classified as vague and none respectively, as indicated by the asterisks (*).

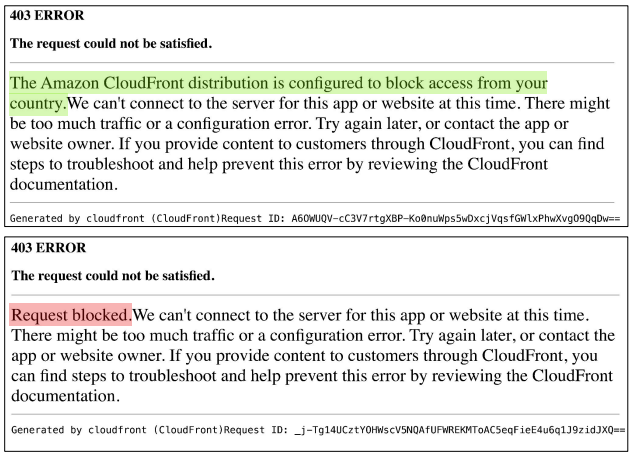

Figure 9: **Interchangeable Blockpages.** We find two fingerprints `Amazon Cloudfront-Geo` and `Amazon CloudFront-Unspecified` are identical other than the highlighted phrases.

while 32.4% of geoblocking domains in the Top 30 categories disclose embargo compliance on their website, we find higher rates within the subset of Technology domains (44.9%) compared to Economy & Finance (25.6%). We acknowledge the difference in sanctions compliance infrastructure surrounding services in these categories; for example, banking institutions have additional mechanisms to prevent unauthorized account registration and more complex and granular terms of service than corresponding Technology domains. Notably, we find categories with much lower levels of disclosures, such as Politics & Advocacy (5.8%).

6.2 Geoblocking Transparency

We observe 88% of geoblocked domains do not serve informative notice of why they are blocked. To further explore how geoblocking impacts users, we focus on the transparency of user-facing HTTP(S) responses, with examples in Table 2.

As discussed in Sec. 5.2.3, we identify 100 geoblocking fingerprints corresponding to HTTP(S) responses retrieved in our measurements. Here, we report the number of domains that correspond to these fingerprints. We find of the 395 domains that geoblock via a response page, 20% (79) of domains provide informative, explicit geoblocking disclosures to the user (*e.g.* "This Service is not availible [sic] in Cuba.") Another 46% (182*) of domains provide vague user-facing information, stating some form of blocking or denial-of-service (*e.g.* `AccessDenied`). The final 34% (135*) of domains either provide no information, a blank page, plain error code, or notice of unsuccessful requests. Interestingly, we find higher levels of disclosures in domains serving informative blockpages (59.5%) than in those serving vague (30.2%) or no information (30.4%).

Still, we emphasize that 88% (480) of geoblocked domains of 546 do not serve informative notice of why they are blocked. These websites demonstrate a clear lack of effort regarding end-user transparency; for example, `match.com`, which denies access via a TCP timeout, discloses in its terms of service, visible only to non-embargoed users, "[users] are not located in a country that is subject to a U.S. Government embargo." This is especially harmful to end users in Cuba; since they cannot see this disclosure and are met with an uninformative TCP error, they are unable to determine whether their failed request was subject to geoblocking or the result of poor network conditions.

Domain-Specific Blockpages. *We find 39 domains serving informative or vague non-standard blockpages with non 403 or 451 status codes.* Throughout our fingerprint identification and clustering process, we identified a specific set of blockpages with non-standard presentation. Recall in Sec. 5.2.3 we find these responses by filtering for content differences compared to the control and subsequently identify them as relevant blockpages via their use of language indicative of geoblocking. As an example, `etsy.com` returns a 200 OK response code in our measurements, suggesting a page was properly returned by the server. However, upon inspection, we see that the page is complete with all background artifacts including headers and links, but presents a sanctions policy page detailing their compliance policy. While this provides increased transparency to the end user, it presents a unique

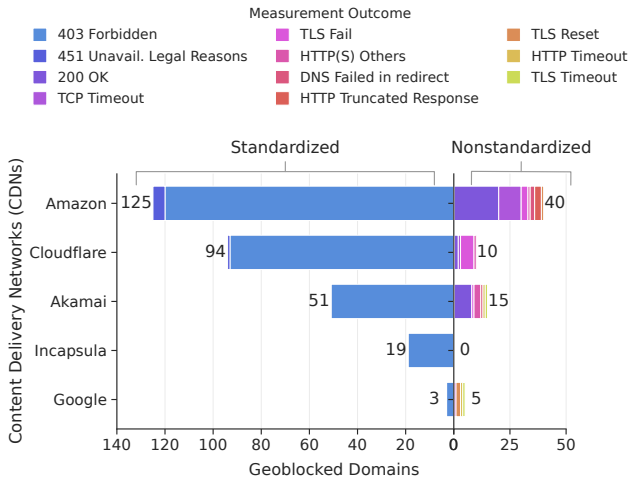

Figure 10: **CDN vs. Geoblocked Domains.** We show the number of geoblocked domains by CDNs and highlight relevant measurement outcomes. Note, 403 or 451 status code responses are on the left, indicating adherence to a standard protocol; all other implementations are on the right.

challenge in identifying geoblocking and a key auditability issue in the geoblocking ecosystem.

Small-scale User Study. Recall in Sec. 4 that we collected 36 domains from respondents and a further 95 from the repository we were pointed to. We find 37% (46) of these 123 domains experience geoblocking: one in the DNS stage, seven in the TCP stage, two in the HTTP stage, and 36 that serve HTTP(S) blockpages, of which six are informative and 10 are vague; the rest provide no additional information. We also explore our category-wise results in relation to our small-scale user study. Our findings corroborate our respondents' identification of the E-commerce and Communication Tools categories as containing high levels of geoblocking; we find 39 and 17 domains, respectively. Note that Communication Tools is an aggregated category, as denoted in Fig. 8.

6.3 CDN Geoblocking Standardization

We find domains hosted on five out of six prominent CDNs (e.g. Amazon, Cloudflare, Akamai) deploy a wide variety of geoblocking implementations, with 7.7% of CDN hosted domains serving blockpages with a 200 OK response. Although CDNs can streamline geoblocking for service providers by standardizing ways to serve blockpages, we still find that geoblocking implementations vary widely, even within individual ASes associated with CDNs, as shown in Fig. 10. Only Incapsula (AS 19551) serves uniform fingerprints with 403 responses; 17 domains serve the `Incapsula incident` fingerprint [35] and two use the `AccessDenied` fingerprint.

Across the other CDNs, the variance is greater. For instance, domains hosted in peering Amazon ASes, Amazon-02 (AS 16509) and Amazon-AES (AS 14618), engage in geoblocking through all layers of the network stack; for example, `megaphone.fm`, a podcast publishing service, blocks by sending a RST packet during the TLS connection, while `ecwid.com`, an e-commerce platform, blocks via poisoning the DNS resolution. In another case, domains hosted in peering Cloudflare ASes, CloudflareNet (AS 13335) and Cloudflare London (AS 209242), serve blockpages with not only 403 and 451 status codes but also with 302 and 200 OK. These blockpages have different transparency levels as well. Some are not informative, such as `error code: 1009` and `error code: 1020`, served by 63 domains, which are Cloudflare designated codes indicating denied access; others are more explicit and transparent, including phrases such as "...Sorry you have been blocked" served by 17 domains.

In an interesting observation, we find two fingerprints `Amazon CloudFront-Geo` and `Amazon CloudFront-Unspecified` are identical other than the switching of the phrases "The Amazon CloudFront distribution is configured to block access from your country" and "Request blocked." The serving of both pages in Amazon-02 (AS 16509) emphasizes not only the arbitrary amount of information provided to end users, but also suggests the ease of deploying informative blockpages as opposed to uninformative ones. This is detailed in Fig. 9.

6.4 Current Improvement Strategies

We find 44% of domains in Technology, Information Technology, Business, Content Servers, Communication Tools, and Education categories implement geoblocking, despite offering free-tier service options. The U.S. government has expressed support for a free Cuban Internet and promoted categories and types of services that are not restricted by embargo sanctions, as discussed in Sec. 2.1. Combining this with our manual exploration of free-tier service offerings in Sec. 5.4, we make a best-effort attempt to quantify the efficacy of these current improvement strategies.

We find domains in categories such as Technology (128), Information Technology (27), Business (87), Content Servers (10), Communication Tools (17), and Education (25) are geoblocked, see Fig. 8. In particular, we find geoblocking services such as `mailchimp.com` (Business), an email marketing platform, and `gitlab.com` (Information Technology), an open-source platform for software development, that additionally disclose their compliance with embargo sanctions regardless of the aforementioned provisions. This emphasizes the tendency of service providers to err on the side of caution with regard to compliance.

We find 44% of geoblocking domains in these lower risk categories provide free-tier service offerings. As all geoblocking we identify is enforced at the top-level domain, this indicates a blanket restriction of access to all services provided

by the domain. For example, `adobe.net`, a computer software company, blocks via a TCP timeout despite providing several free online services, such as PDF editors. Another domain, `databricks.com`, a data analytics platform, serves a 403 Forbidden page, but offers a free community version of its cloud-based services.

7 Discussion and Recommendations

Given the lack of standardized geoblocking implementation and transparency, along with the inefficacy of the U.S. current improvement strategies for a free Cuban Internet, in this section, we outline recommendations for improvement.

What can service providers do to help? We believe service providers should push for transparency and assist in reducing Internet fragmentation. While we understand that service providers may prefer a silent de-risking approach, we believe that they have a responsibility to communicate clearly and effectively. We encourage service providers to serve meaningful blockpages that contain explicit language indicative of geoblocking. This change would significantly help user-level transparency and come at little to no cost for providers. Regarding measurement-level transparency, we observe that domains implementing geoblocking unfortunately do not always return a 403 or 451 status code, despite these codes being most directly related to geoblocking. We urge providers to work towards standardization of geoblocking implementation. These proposed improvements would increase the ease of auditing the ecosystem by allowing for greater integration of automation and proportionally reducing time-consuming manual analysis.

Furthermore, we acknowledge blocking at the top-level domain with both free and paid offerings is simpler and lower risk for service providers. However, in the case that they have the means and a vested interest in promoting Internet freedom, strictly-free software could be consigned to a new domain and made accessible even to those under embargo. In support of this, we find that there are no existing cases where a digital service provider was punished by U.S. authorities for providing a free online service in Cuba.

What can policymakers do to help? Our findings highlight two necessary adjustments for policymakers. First, given the stated goal of increasing access to information in Cuba, policymakers must engage with service providers around the world to ensure sanctioned users are not overly geoblocked. Second, apparent attempts to distinguish between free and paid services in regulation have not resulted in equal consideration among service providers — many service providers still geoblock Cuban users from free versions of their services. We admit that there is currently a lack of motivation for service providers to act on our recommendations due to a minimal yield from incurred costs and efforts; we are keenly aware that policymakers are best positioned to motivate changes and must do so to back up their communicated intentions.

There is an irony to our findings. Over the past three fiscal years, the U.S. government has spent over $47M on the Office of Cuba Broadcasting to bring free information to the Cuban people. In 2016, the U.S. Coordinator for International Communications and Information Policy, Ambassador Daniel Sepulveda, pushed the Cuban government to adopt transparent Information and Communication Technology procurement policies, encourage competition, and liberalize the telecommunications sector [50]. Access to information is a clear goal of U.S. policy toward Cuba. However, the U.S. government must modify existing regulations if they intend for Cuban citizens to have access to free and widely available services. Existing attempts to push providers to allow free access as a lower-risk alternative have not resulted in meaningful changes.

Non-Sanctions Geoblocking. In Sec. 5.4, we discuss how we manually examine geoblocking domains to better characterize geoblocking attribution. We systematically evaluate domains for privacy policies, legal notices, and other compliance pages; we additionally document the country of registration, finding non-U.S. registered companies still disclose adherence to U.S. sanctions — underscoring the opaqueness of the compliance ecosystem. However, we acknowledge that aside from sanctions, there are other causes for geoblocking. For example copyright/licensing policies are often country-specific; differences in privacy legislation can impact service availability; and distributed denial-of-service (DDoS) prevention can discriminate against IPs from countries with a history of engaging in DDoS attacks. We note that Cuba has no increased privacy legislation like GDPR, or a history of launching of DDoS attacks like Russia.

Limitations. We acknowledge some limitations of our work. First, we are only able to test domain accessibility over HTTPS, so we cannot study domains available only over HTTP. However, as the majority of Internet traffic is over HTTPS, this is not a major hindrance. Second, due to the difficulties associated with mapping IP addresses to autonomous systems, we use a combination of GeoIP2 Tables and Censys to accurately link IPs to their associated AS [14].

8 Conclusion

In this work, we present the first in-depth and systematic investigation into the effects of sanctions on web content inaccessibility, specifically geoblocking in Cuba. We conduct network measurements on the Tranco Top 10K augmented with additional domains from our small-scale user study, which we further use to characterize the impact of digital sanctions for respondents in Cuba. We identify 546 geoblocking with 17 diverse implementations, ranging from DNS failures to HTTP(S) response pages with a variety of status codes. We highlight high levels of geoblocking in categories such as

Technology and Business, as well as identify the most common implementations like web servers dropping TCP SYN packets or serving blockpages with 403 status codes. Additionally, we discover a lack of user-facing transparency, finding that 88% of geoblocked domains do not serve informative notice of why they are blocked. We also highlight a lack of measurement-level transparency among HTTP(S) blockpages, identifying 32 instances of blockpages served with 200 OK status codes. Finally, we note the inefficacy of current improvement strategies and we provide recommendations to both policymakers and service providers to reduce Internet fragmentation. We hope our work inspires future interdisciplinary studies on the real-world impacts of embargo sanctions on geoblocking.

Acknowledgments

The authors thank the anonymous reviewers for their helpful feedback. We are also deeply grateful to Doug Madory, Afsah Anwar, Jedidiah Crandall, Kyle Astroth, and Yusei Uehara for all their help and support on this work. This research was supported by National Science Foundation grants CNS-2141512 and CNS-2237552.

References

[1] A&E Talk Politics. The U.S.-Cuban Embargo. https://open.spotify.com/episode/31TLJ731IrEEPiE2sQqEve, 2022.

[2] Sadia Afroz, Michael Carl Tschantz, Shaarif Sajid, Shoaib Asif Qazi, Mobin Javed, and Vern Paxson. Exploring server-side blocking of regions. *arXiv preprint arXiv:1805.11606*, 2018.

[3] Alborada. Cuban Socialism vs US Sanctions w/ Helen Yaffe. https://open.spotify.com/episode/7CbMsvD775i5mcxhTi5UzP, 2023.

[4] Asamblea Nacional del Poder Popular. Decreto no. 209/96, Sobre el acceso de la República de Cuba a Redes de Alcance Global. http://ordiecole.com/cuba/209-1996.pdf, 1996.

[5] Farzaneh Badiei. Sanctions and the Internet. Technical report, RIPE Labs, 2023.

[6] Michael Bailey, David Dittrich, Erin Kenneally, and Doug Maughan. The menlo report. *IEEE Security & Privacy*, 2012.

[7] David A. Baldwin. *Economic Statecraft*. Princeton University Press, Princeton, N.J., 1985.

[8] John B Bellinger, Thomas A Shannon, John P Barker, Baruch Weiss, and Sean A Mirski. Two years of title III: Helms-Burton lawsuits continue to face legal obstacles: Advisories, 2021.

[9] Vera Bergengruen. US sanctions are blocking Cuban activists from online platforms, Nov 2021.

[10] Abhishek Bhaskar and Paul Pearce. Many roads lead to Rome: How packet headers influence DNS censorship measurement. In *USENIX Security Symposium*, August 2022.

[11] Zachary S. Bischof, John P. Rula, and Fabián E. Bustamante. In and out of Cuba: Characterizing Cuba's connectivity. In *Internet Measurement Conference*, 2015.

[12] Natalie Campbell and Nermine El Saadany. Sanctions Can Deny Internet Access When People Need It Most, September 2022.

[13] Censored Planet. GeoInspector - Geoblocking Measurement Toolkit. https://github.com/censoredplanet/geoinspector, 2023.

[14] Censys. Exposure management and threat hunting solutions. https://censys.com/, 2024.

[15] CNET. Companies that have left Russia: The list across tech, entertainment, finance, sports. https://cnet.com/news/politics/companies-that-have-left-russia-the-list-across-tech-entertainment-finance-sports/, June 2022.

[16] Cuban Liberty and Democratic Solidarity (LIBERTAD) Act of 1996 (H.R.927), 1996.

[17] Council on Foreign Relations. Russia's internet censor is also a surveillance machine. https://cfr.org/blog/russias-internet-censor-also-surveillance-machine, Sept 2022.

[18] Cuban Opensourcers. Awesome list about tech sites restricted for cuba. https://github.com/cuban-opensourcers/cuban-restricted, 2023.

[19] Bryan R. Early and Timothy M. Peterson. The enforcement of U.S. economic sanctions and global de-risking behavior. *Journal of Peace Research*, 2023.

[20] ETECSA. Empresa de telecomunicaciones de cuba s.a. https://x.com/etecsa_cuba, 2014.

[21] Marc Frank. Cuba lifts ban on computer and DVD player sales. https://reuters.com/article/us-cuba-reforms/cuba-lifts-ban-on-computer-and-dvd-player-sales-idUSN1329909720080313, Mar 2008.

[22] Freedom House. Turkey: Passage of social media law curtails human rights online. https://freedomhouse.org/article/turkey-passage-social-media-law-curtails-human-rights-online, July 2020.

[23] Freedom House. Cuba: Freedom on the Net 2022 Country Report. https://freedomhouse.org/country/cuba/freedom-net/2022, 2022.

[24] Johan Galtung. On the Effects of International Economic Sanctions: With Examples from the Case of Rhodesia. *World Politics*, 19(3):378–416, 1967.

[25] Guerrilla History. The History of US Sanctions on Cuba w/ Helen Yaffe. https://open.spotify.com/episode/0tGOPciHKZqtmSNwMDhmG6, 2022.

[26] Bruce W. Jentleson. *Sanctions: What Everyone Needs to Know®*. What Everyone Needs To Know®. Oxford University Press, Oxford, New York, September 2022.

[27] Ben Jones, Roya Ensafi, Nick Feamster, Vern Paxson, and Nick Weaver. Ethical concerns for censorship measurement. In *ACM SIGCOMM Workshop on Ethics in Networked Systems Research*, Aug 2015.

[28] Ben Jones, Tzu-Wen Lee, Nick Feamster, and Phillipa Gill. Automated detection and fingerprinting of censorship block pages. In *Internet Measurement Conference*, 2014.

[29] Sam P. Kellogg. Blocked ports: Sanctions and software in networked. *Media Fields Journal - Issue 17*, Aug 2022.

[30] Renuka Kumar, Apurva Virkud, Ram Sundara Raman, Atul Prakash, and Roya Ensafi. A large-scale investigation into geodifferences in mobile apps. In *USENIX Security Symposium*, 2022.

[31] Victor Le Pochat, Wouter Joosen, Maciej Korczyński, Samaneh Tajalizadehkhoob, and Tom Van Goetham. Tranco: A Research-Oriented Top Sites Ranking Hardened Against Manipulation, 2023.

[32] Chao Liu and Ernesto Falcon. There is Nothing Fair About the European Commission's "Fair Share" Proposal, 2023.

[33] Nadia L Luhr. Iran, Social Media, and U.S. Trade Sanctions: The First Amendment Implications of U.S. Foreign Policy. *First Amendment Law Review*, 8(2), 2010.

[34] Sarah Marsh. Cuba launching internet on cellphones. https://reuters.com/article/us-cuba-internet/cuba-launching-internet-on-cellphones-idUSKBN1O40B5, Dec 2018.

[35] Allison McDonald, Matthew Bernhard, Luke Valenta, Benjamin VanderSloot, Will Scott, Nick Sullivan, J. Alex Halderman, and Roya Ensafi. 403 forbidden: A global view of CDN geoblocking. In *Internet Measurement Conference*, 2018.

[36] Abraham Mhaidli, Selin Fidan, An Doan, Gina Herakovic, Mukund Srinath, Lee Matheson, Shomir Wilson, and Florian Schaub. Researchers' experiences in analyzing privacy policies: Challenges and opportunities. *Proceedings on Privacy Enhancing Technologies*, 2023.

[37] Giovane C. M. Moura, John Heidemann, Ricardo de O. Schmidt, and Wes Hardaker. Cache me if you can: Effects of DNS time-to-live. In *Internet Measurement Conference*, 2019.

[38] Nicholas Mulder. *The Economic Weapon: The Rise of Sanctions as a Tool of Modern War*. Yale University Press, New Haven London, 2022.

[39] Arvind Narayanan and Bendert Zevenbergen. No encore for Encore? Ethical questions for web-based censorship measurement, Sept. 2015.

[40] OFAC. Fact sheet: Preserving agricultural trade, access to communication, and other support to those impacted by Russia's war against Ukraine. https://ofac.treasury.gov/media/922206/download?inline, Apr 2022.

[41] Office of the Federal Register. Electronic code of federal regulations (e-CFR). https://ecfr.gov/, 2023.

[42] Isabella Oliver and Mariakarla Nodarse Venancio. Understanding the failure of the U.S. embargo on Cuba. https://wola.org/analysis/understanding-failure-of-us-cuba-embargo/, 2022.

[43] OONI. Open observatory of network interference. https://ooni.org/, 2024.

[44] Craig Partridge and Mark Allman. Addressing ethical considerations in network measurement papers. In *ACM SIGCOMM Workshop on Ethics in Networked Systems Research*, 2015.

[45] Paul Pearce, Roya Ensafi, Frank Li, Nick Feamster, and Vern Paxson. Augur: Internet-wide detection of connectivity disruptions. In *IEEE Symposium on Security and Privacy*, 2017.

[46] Gerhard Peters and John T. Woolley. Proclamation 3447—embargo on all trade with Cuba. https://presidency.ucsb.edu/node/237154, Feb 1962.

[47] Ram Sundara Raman, Mona Wang, Jakub Dalek, Jonathan Mayer, and Roya Ensafi. Network measurement methods for locating and examining censorship devices. In *International Conference on Emerging Networking EXperiments and Technologies*, 2022.

[48] Reethika Ramesh, Ram Sundara Raman, Matthew Bernhard, Victor Ongkowijaya, Leonid Evdokimov, Anne Edmundson, Steven Sprecher, Muhammad Ikram, and Roya Ensafi. Decentralized Control: A Case Study of Russia. In *Network and Distributed System Security*. The Internet Society, 2020.

[49] Reethika Ramesh, Ram Sundara Raman, Apurva Virkud, Alexandra Dirksen, Armin Huremagic, David Fifield, Dirk Rodenburg, Rod Hynes, Doug Madory, and Roya Ensafi. Network responses to Russia's invasion of Ukraine in 2022: A cautionary tale for internet freedom. In *USENIX Security Symposium*, 2023.

[50] Daniel Sepulveda. Cuba and the Internet: Choices, Challenges, and Opportunities, January 2016.

[51] Jarred O Taylor. Information Wants to be Free (of Sanctions): Why the President Cannot Prohibit Foreign Access to Social Media Under U.S. Export Regulations. *William & Mary Law Review*, 54(1), 2012.

[52] Michael Carl Tschantz, Sadia Afroz, Shaarif Sajid, Shoaib Asif Qazi, Mobin Javed, and Vern Paxson. A bestiary of blocking: The motivations and modes behind website unavailability. In *USENIX Workshop on Free and Open Communications on the Internet (FOCI)*, 2018.

[53] U.S. Department of State. Advancing the free flow of information for the iranian people. https://state.gov/advancing-the-free-flow-of-information-for-the-iranian-people/, September 2022.

[54] U.S. Department of the Treasury. Fact sheet: Supporting the Cuban people's right to seek, receive, and impart information through safe and secure access to the internet. https://ofac.treasury.gov/media/912206/download?inline, 2021.

[55] U.S. Department of the Treasury. De-Risking Strategy. https://home.treasury.gov/system/files/136/Treasury_AMLA_23_508.pdf, 2023.

[56] Richard Verma. Letter and Report for the Congressional Record, December 2009.

[57] Diwen Xue, Benjamin Mixon-Baca, ValdikSS, Anna Ablove, Beau Kujath, Jedidiah R. Crandall, and Roya Ensafi. TSPU: Russia's decentralized censorship system. In *Internet Measurement Conference*, 2022.

[58] Ahmad Yulianto and Rina Supriatnaningsih. Google Translate vs. DeepL: a quantitative evaluation of close-language pair translation. *AJELP: Asian Journal of English Language and Pedagogy*, 9(2), 2021.

[59] Bendert Zevenbergen, Brent Mittelstadt, Carissa Véliz, Christian Detweiler, Corinne Cath, Julian Savulescu, and Meredith Whittaker. Philosophy meets Internet engineering: Ethics in networked systems research. GTC Workshop Outcomes Paper, 2015.

A Appendix

A.1 User Questionnaire

We provide the full questions that we ask our respondents from Sec. 4. We first present the respondent with the statement detailing the potential risk and the measures that we take to protect their anonymity and privacy, and then ask the following questions:

Disclosure of Risk. In this project, we are trying to understand how sanctions on different countries affect access to the Internet for users in that country. Anecdotal evidence and previous studies have shown that many online services tend to restrict access to users from certain regions. Please read the following information and consent document for all details regarding this study. Your participation and responses are completely anonymous. By proceeding with this survey, you express your consent and agree to be a part of this study. You may take this survey if you have faced restrictions by Internet services. Please keep in mind that this survey is intended to collect data about restrictions by content providers, and not by your ISP and Government. We will be collecting the answers you provide as a survey participant, but note that we do not collect your name, email address or any personal identifier along with the survey. Hence, we will not be able to attribute a survey response to any one respondent. The survey questions merely focus on our research concerns and do not ask for any personally identifiable information. We will only receive your survey response once you hit the submit button at the end of the survey. Please know you can always come back and update your answers. Most questions are not required, and we value any information you provide. Thank you very much for taking part in our user survey investigating the impact of Sanctions on Internet users.

Demographics. *Demographic information.*

Q1-3. Age, Country of Residence, Occupation

Understanding the websites that are inaccessible, and the impact of their unavailability. *In this section, we ask questions about the websites and services that are generally inaccessible in your country, and about the impact that this unavailability and inaccessibility has had on you as a user.*

Q4. Can you please list some other websites, desktop applications, and mobile applications that impose server-side blocking towards users in your country?

Q4.1. Can you please tell us what you see when you try to access the above services, apps, and websites? (Choose all) ☐ Access Restricted Page ☐ Browser or App Error ☐ Not available for users in your region ☐ Nothing (no change) ☐ Blank page ☐ Apps not found in stores ☐ No download button ☐ Other

Q4.2. If possible, could you please upload some example screenshots of the response you receive when you visit restricted sites?

Q5. What categories of services (website, applications, mobile apps) do they generally belong to? (Choose all)

☐ Alcohol & Drugs ☐ Anonymization and circumvention tools ☐ Communication tools ☐ Culture ☐ E-commerce ☐ Economics ☐ Environment ☐ File-sharing ☐ Gambling ☐ Gaming ☐ Government ☐ Hacking tools ☐ Hate speech ☐ Hosting and blogging platforms ☐ Human rights issues ☐ Intergovernmental Organizations ☐ LGBT ☐ Media sharing ☐ Miscellaneous content ☐ News media ☐ Online dating ☐ Political criticism ☐ Pornography ☐ Provocative attire ☐ Public health ☐ Religion ☐ Search engines ☐ Sex education ☐ Social networking ☐ Terrorism and militants ☐ Other

Q6. The inaccessibility of the above services affects the following aspects of my day-to-day life:

○ Online learning ○ Education ○ Career opportunities ○ Work obligations ○ Staying up to date ○ Financial well being ○ Mental health ○ Time overhead ○ Loss of entertainment

Q7. Please tell us why you chose the above options (or others), and elaborate the impact that the inaccessibility of the services had on your personal and professional life.

Understanding users' handling of the inaccessibility. *In this section, we ask questions regarding how you handle and manage the inaccessibility of necessary websites and services.*

Q8. If you use services to bypass restrictions, where do you find information about these services?

Q9. If you are comfortable disclosing, how do you bypass these restrictions to reach the website or service? (Choose as many as you like)

○ VPNs ○ Tor Browser ○ Proxy services ○ Other

Q10. Are there any services available in your country that act as an alternative to the unavailable services? How do they compare to the unavailable services?

Q10.1. How safe do you feel using the alternative services?

Q11. Is there anything related to website unavailability and inaccessibility that you wish to share with us?

A.2 Supplementary Figures & Tables

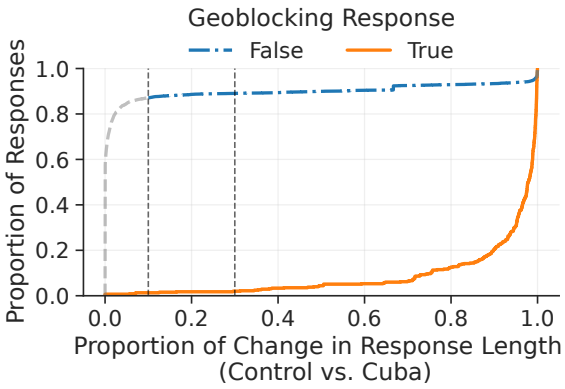

Figure 11: **Proportion of Change in Response Length.** We compare the response lengths between control and Cuba measurements. We annotate this with values (0.1 and 0.3) that denote our labeling thresholds as discussed in Sec. 5.2.3.

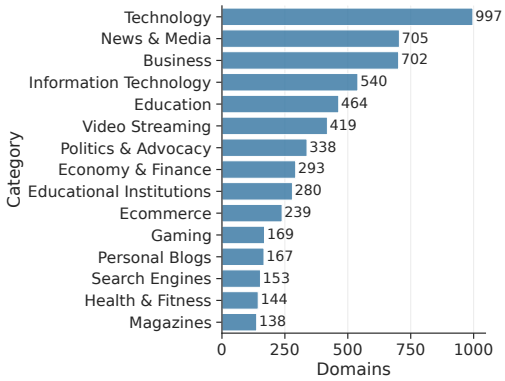

Figure 12: **(Passed Control) Top 15 Categories and Domains.** We analyze the category breakdown of all domains passing control measurements as detailed in Sec. 5.1.

Status Codes	Description
200 OK	Request succeeded
201 Created	Request succeeded, and new resource was created
301 Moved Permanently	Resource moved permanently, new URL given
403 Forbidden	Client is not authorized to view the resource
404 Not Found	Server cannot find the resource
451 Unavailable for Legal Reasons	Client requested a resource that the server cannot provide access to because of legal reasons
503 Service Unavailable	Server cannot handle the request due to maintenance or overload

Table 3: **HTTP Response Status Code.** Summary of HTTP status codes that we encountered. This complements our measurement findings in Table 1.