Free and Open
COMMUNICATIONS
https://foci.community on the Internet

CensorWatch: On the Implementation of Online Censorship in India

Divyank Katira Gurshabad Grover Kushagra Singh Varun Bansal
Centre for Internet and Society

Free and Open Communications on the Internet 2023(1), 35-45

CensorWatch: On the Implementation of Online Censorship in India

Divyank Katira Gurshabad Grover Kushagra Singh Varun Bansal
Centre for Internet and Society

Abstract

State authorities in India order domestic internet service providers (ISPs) to block access to websites and services. We developed a mobile application, CensorWatch, that runs network tests to study inconsistencies in how ISPs conduct censorship. We analyse the censorship of 10,372 sites, with measurements collected across 71 networks from 25 states in the country. We find that ISPs in India rely on different methods of censorship with larger ISPs utilizing methods that are harder to circumvent. By comparing blocklists and contextualising them with specific legal orders, we find concrete evidence that ISPs in India are blocking different websites and engaging in arbitrary blocking, in violation of Indian law.

1 Introduction

Nation states routinely engage in online censorship to control information flows and restrict citizens' access to information that governments view as unlawful or undesirable. The Government of India exercises such authority in forms varying from internet shutdowns to blocking of specific online resources. For the latter, India follows a 'decentralized' model of information controls. Unlike countries like Iran and China where the states directly control most infrastructure for web censorship, [1, 24] state authorities in India are empowered through law to order internet service providers (ISPs) that operate in the country to block certain websites for their customers.

There are several peculiar impediments to studying web censorship in India. First, content blocking orders are confidential, [20, 21, 16] making it difficult for researchers or citizens to ascertain what websites are supposed to be blocked in India. [9] This is unlike Russia, for instance, which also follows a semi-decentralized system of censorship, [15, 25] but where the telecommunications regulator maintains a public list of all websites blocked by law. [10]

Second, Indian law does not mandate ISPs to follow any specific technical method of blocking websites or URLs. [22] With 66 ISPs [1] [11] being free to choose any blocking method they wish, any study must be run from several networks and locations to paint an accurate picture of online censorship in India.

With this in mind, we developed a mobile application, CensorWatch, which allowed any Indian internet user with an Android device to run network tests that determine whether a set of websites is blocked by their ISP. This paper describes the CensorWatch network tests, and analyzes measurement data for 10,372 websites contributed by 331 users from 25 states in India, across 71 autonomous systems (ASes). We make the following notable contributions:

- **Largest set of potentially blocked websites**: To create a list of potentially blocked websites, we performed an extensive survey of previous scholarly work, public and leaked government orders, court orders in the public domain or obtained through right to information requests, and user reports. Finally, we end up with 10,372 unique hostnames. This is almost twice as large as the set collected by prior work, Singh, et al. [22].

- **AS coverage**: While previous studies have only reported on the censorship mechanisms from 6-9 ASes, [22, 26], we collected and analyzed measurements from 71 ASes across the country. A wide coverage across networks allows us to get a sense of how smaller ISPs outside of large cities in India conduct web censorship.

- **Distributed probes**: Prior studies were also limited to measurements from a handful of locations. Our data set spans measurements from 25 states in the country, allowing us to study variations in blocking by region.

- **Contextualisation with legal orders and developments**: Our work is the first to contextualize blocking of specific websites with relevant legal orders. We create

[1] Our paper's analysis is at the level of Autonomous Systems (ASes); a single ISP can operate multiple ASes.

specific evidence of ISPs being non-compliant with legal orders, and arbitrarily blocking websites and services without a legal basis.

The rest of the paper is structured as follows. We first summarize the related work in the field. We then describe our methodology: this includes our network tests that are capable of measuring DNS-, HTTP-, and SNI-based forms of censorship. We also detail our data cleaning and analysis process. Finally, we summarize our results, which include demonstrating variances in blocklists across ISPs, and identifying whether censorship varies across regions in India. We then contextualize website blocking with legal orders that we were able to access. Finally, we give actionable recommendations to the government, regulators and ISPs in India.

2 Related work

There are three studies of web censorship in India that rely on direct measurement. Singh, et al. [22] and Yadav, et al. [26] both rely on manual measurement from a single vantage point each for a limited number of ISPs. They cover the large and popular ISPs, but fail to describe the censorship techniques used by smaller ISPs, and are incapable of commenting on whether there are regional differences in online censorship in India.

Global censorship measurement platforms can generally be used for our purposes, since they permit testing from any network. For instance, a part of Chinmayi SK's report relies on running Open Observatory of Network Interference (OONI) network tests from five ISPs, but is limited to the state of Manipur. [23] The test lists used by OONI are quite limited in size,[2] and OONI currently does not have the ability to associate any geolocation data with measurements. OONI would have been unsuitable for our purposes because of our large set of potentially blocked websites and our intention to study regional trends (within India) in censorship.

CensoredPlanet runs remote measurements (as opposed to direct measurements), i.e. measuring censorship by contacting open servers in the country through probes in other countries. This can greatly mitigate the risks posed to users, but assumes that middleboxes responsible for censorship are monitoring traffic bidirectionally, which can result in under-reporting of censorship. Our preliminary tests also indicated that some websites that CensoredPlanet reported as being blocked in India [14] were in reality accessible through multiple ISPs. We also could not assume that smaller ISPs in India would host open servers.

Due to the limitations of these tools and previous work, which would prevent us from studying the blocking of a large set of potentially blocked websites and commenting on regional trends, we decided to develop a mobile application – CensorWatch. We published and distributed this app on the Google Play Store for Android devices,[3] and publicized it through our personal, professional and social media networks.

3 Legal and ethical considerations

Our methodology relies on Indian residents running direct measurements that detect whether a particular website is blocked on their ISP. To a passive observer, the network traffic of these tests can seem to mimic that of a person trying to access a blocked website. This can pose legal or extralegal risks for those engaging in such measurements, especially for activists or those already under surveillance.

Our research could not uncover any instance of an Indian resident facing legal or extralegal action because of their attempts to access censored material online. Sections 79 and 69A of the IT Act, and the 'Blocking Rules' are used to order ISPs to block websites in India: none of these regulations prohibit users from trying to access blocked content in general.[20, 21, 16]

Many of the websites we test for contain ostensibly copyright-infringing content, pornography, politically sensitive content, or may host content that we have not manually inspected or analysed. In the context of copyright infringement, the Madras High Court specifically stated that simply attempting to access a 'blocked' website is not illegal in itself.[2]

Sections 67, 67A and 67B can be used to target anyone who publishes, transmits or "causes to be transmitted" obscene material, sexually explicit material, and child sexual abuse material respectively. [17, 18, 19] Therefore, we design our network tests so that they do not access any actual content from websites in our list. Specifically,

- *DNS*: CensorWatch queries the network-assigned DNS resolver for a given hostname. The actual website is not contacted for this test.
- *HTTP*: CensorWatch only retrieves the robots.txt file from the server associated with a specific hostname.
- *SNI*: Our test to detect SNI-based blocking attempts to connect to a known TLS server unassociated with the hostname.

We seek the informed consent of each CensorWatch user before any network tests are run by the app. We also do not retain any personally identifiable information of users: the app requires no special permissions, and the IP addresses of our users are deleted in daily batches from our database (after a script associates it with the AS).

[2] While OONI Run does allow users to run tests on a custom list of websites, it is limited in the sense that it relies on custom URLs. Our preliminary tests indicated that it would take multiple hours for a single device to run the web connectivity test on 10,000 URLs.

[3] Other mobile operating systems were not considered as Android accounts for more than 95% of smartphone users in India.

4 Methodology

4.1 List of potentially blocked websites

Website blocking orders are issued by the Government of India under sections 69A and 79 of the Information Technology (IT) Act. [20, 21] Orders issued under section 69A are governed by a confidential procedure, [16] and it remains difficult to access specific orders or a complete list of blocked websites in India. Courts may also issue website blocking orders, in which case they are generally accessible to the public.

Singh, et al. [22] compiled a list of potentially blocked websites from government orders, court orders, and public reports. We update and expand on the same list. The biggest addition to the list is a corpus leaked by a whistleblower to the Internet Freedom Foundation. [6] We also added websites to the list based on media and user reports about specific or suspected government action.

Our final list comprises 10,372 websites, which is around double of the list collected by Singh, et al. [22] Most of our final set of the potentially blocked websites is derived from court orders, and largely relates to copyright-infringement. A minority of these relate to pornography. The executive-issued orders relate to human rights, domestic and international politics. The portion of our list contributed by users (collected by the Internet Freedom Foundation), contains websites related to messaging, censorship circumvention, and piracy.

4.2 System architecture

Our system consists of four entities: (1) the CensorWatch server, (2) a control server that gets uncensored responses for websites in our list, (3) a known server used by our SNI test, and (4) our network of test nodes, i.e. Android devices that have installed the CensorWatch app and run network tests.

4.2.1 CensorWatch server

The CensorWatch server acts as the primary coordinating entity for all the test nodes. We set up a script on the CensorWatch server that runs daily to resolve the DNS name for each website in our list using a non-censorious public DNS over HTTPS server. When a CensorWatch user initiates testing on their device, the CensorWatch server serves the list of the domain names that are resolved successfully. The CensorWatch server receives measurements from users' devices, and regularly runs analysis scripts that detect DNS, HTTP, and SNI -based censorship (described under section 4.3) on all received observations, and saves the results.

4.2.2 Control server

We set up a control server, a DigitalOcean instance in California, United States. This control server retrieves the list of websites from the CensorWatch server, and attempts a HTTP connection to each website in the list (described in section 4.3.2). It saves each response.

4.2.3 Known TLS server

We identify a TLS server that is set up to respond to all attempted TLS connections, even if it does not host the website provided in the Server Name Indication (SNI) in the ClientHello.

4.2.4 Network of test nodes

Our Android app CensorWatch runs network tests (described under section 4.3) that help determine whether a given hostname is blocked on the device's network connection. After installing the application, the users are guided through the aims of the research project, the privacy policy, the possible risks of running such network measurements, and the estimated time for running the tests. After reading this information, the users self-report which state in India they are running the test from (to help us identify regional trends in censorship) and then initiate a run of the network tests.

At the beginning of each 'run', the Android app retrieves the list of potentially blocked websites from the CensorWatch server. Each 'run' consists of running the following tests on each hostname in the list: (1) the DNS test, (2) the HTTP test, and (3) the SNI test. The result of each test (what we call a 'measurement') is sent to the CensorWatch server.

4.3 Network tests

4.3.1 The DNS test

The Domain Name System (DNS) is responsible for resolving human-readable domain names (like example.com) to their IP address(es). Traditionally, DNS queries have been carried over plaintext, and are vulnerable to interception and forging. While encrypted DNS protocols like DNS over HTTPS (DoH) and DNS over TLS (DoT) have been deployed, intercepting plaintext DNS remains a popular way for ISPs to block access to specific websites and services.

For a hostname that they would like to block access to, ISPs can either maintain a false IP address associated with it on their DNS servers (i.e. DNS poisoning), or intercept plaintext DNS queries over the wire and send back a false response to the user device (DNS injection). The 'DNS test' is our methodology to determine whether a given hostname is blocked on an ISP using either of these techniques.

For a given hostname, our control server (based in California) first queries a non-censorious DNS server. We record the response as `true_ip_list`.

A test node (Android device) queries the configured resolver for the DNS record of the website. The device receives a DNS record. The app keeps the first IP address in this record, `resolved_ip`, and discards the others. The app sends

this resolved_ip, the IP address of the device (source_ip) and the device's configured DNS servers to the CensorWatch server for futher analysis. Our initial heuristic for determining whether a hostname is blocked on a particular ISP is summarized in Algorithm 1.

Algorithm 1: Initial heuristic for the DNS test

Result: Determines DNS tampering

```
if true_ip_list = ∅ ∧ resolved_ip = ∅ then
    return censorship = False
if true_ip_list ≠ ∅ ∧ resolved_ip = ∅ then
    return censorship = True
if true_ip_list = ∅ ∧ resolved_ip ≠ ∅ then
    return censorship = True
if resolved_ip ∈ true_ip_list then
    return censorship = False
if resolved_ip.ASN ∈ true_ip_list.ASNs then
    return censorship = False
if resolved_ip.ASN = source_ip.ASN then
    return censorship = True
return censorship = False
```

Note that we take a conservative approach here: if none of the conditions listed are met, we mark the website as not affected by DNS tampering.

After deploying the initial heuristic, we also refine our results to rule out false positives. Specifically, the resolved_ip_list ASN could also match the source_ip ASN in case the website used a CDN and/or the ISP cached the website. To rule out these false positives, we compiled the most common IP address received in response to the DNS queries. This heuristic helps to identify the IP addresses which censorious DNS servers give to users. This approach is similar to Singh, et al [22], and we mark all measurements that encountered that IP address as symptomatic of censorship. We were able to confirm 89% of the suspected blocks in this way.

4.3.2 The HTTP test

The Hypertext Transfer Protocol (HTTP) is commonly used to transfer information and media over the web. If not used with an encrypted transport protocol, plaintext HTTP traffic remains vulnerable to monitoring. Elements in the HTTP flow, such as the HOST header field, can reveal to the ISP the identity of the website any internet user is trying to access. If an ISP intends to block a given hostname, it can detect this information and then disrupt or drop the connection, or send a forged working response to the user.

The 'HTTP test' is our methodology to detect whether a given hostname is blocked using these censorship methods that rely on information in plaintext HTTP. The HTTP test takes a two tuple as an input, a hostname and its IP address resolved via our control server (where we expect a non-censored response). [4]

For each hostname, the control server attempts to fetch the {hostname}/robots.txt, [5] which serves as the uncensored ground truth (it is unlikely that different robots.txt files are served in different regions). We save this (control_response).

For each input tuple, the user device attempts to establish a connection to port 80 with the resolved IP address while specifying the hostname in the HOST header. After establishing the connection, the device makes a HTTP GET request for /robots.txt. We send the received test_response (with the timestamp, headers, code, length, body) to the CensorWatch server.

We conclude that a hostname is censored if either of these conditions is true: (1) The connection attempt results in a connection reset (TCP RST), which is indicative of active network interference and commonly used for censorship. [13] (2) The control_response and the test_response do not match.

To confirm the results of our analysis, we collected the distinct blockpages in our dataset and matched them with each test_response that was marked censored. We found that all of them matched a known blockpage, which confirmed our findings.

4.3.3 The SNI test

HTTPS (HTTP Secure) uses Transport Layer Security (TLS) to achieve confidentiality, integrity and authenticity of network data transfer. A passive network observer can still glean information about the endpoint that a user is trying to connect to from the Server name Indication (SNI) field, an extension in the ClientHello message that carries the name of the website in plaintext.

By monitoring the SNI, ISPs can detect when a user is trying to connect to a website they would like to block, and drop the connection or terminate it by sending a TCP reset.

Our 'SNI test' follows the methodology by Singh, et al [22] to detect if a given hostname is blocked by a network by intercepting TLS connections. For each hostname, a user device attempts to establish five TLS connections to our 'Known TLS server', specifying the given hostname in the SNI field. As noted before, this known TLS server is set up in a way to ignore the SNI and accept all connections. As our SNI test does not depend TLS version used, the client is allowed to negotiate any of the current versions i.e. TLS 1.2 and 1.3. The measurement is recorded and the timestamp, TLS metadata

[4] Note that the CensorWatch server filters out the websites that IP addresses that do not support connections on port 80, and CensorWatch users' devices do not test for connectivity to those hostnames.

[5] A robots.txt file defines how crawlers will access material on the website. See https://www.rfc-editor.org/rfc/rfc9309.txt

(protocol version, cipher suite, etc.), and the number of successful and unsuccessful TCP/TLS connections is sent to the CensorWatch server.

We call a website censored if all attempts to connect to the hostname result in Socket exceptions (which occurs when the remote port is reachable, but a connection could not be established with it).

4.4 Analyzing ISP blocklists and blocking methods

Across the three tests (DNS, HTTP and SNI), we collected 8.9 million measurements from 331 runs. This subsection describes how we went about cleaning the dataset, and classifying website blocks and censorship techniques used by ISPs.

4.4.1 Removing invalid measurements

From the user's Internet Protocol (IP) address, we were able to determine which network the tests were run from. As we are only looking to study censorship on public networks based in India, we use information from the `ipinfo.io` database to exclude all measurements that came from IP addresses outside the country; or those belonging to known Virtual Private Networks (VPNs), the Onion (Tor) network, proxy services, and hosting services.

4.4.2 Determining region

To understand regional differences in blocking, we needed to determine the state from which tests were being run. We had two ways to do this – geolocate users using their IP address or rely on user-reported regions collected in the mobile application. We found that in 98 out of the 331 test runs, the region reported by the user did not match that of the geolocation derived from the IP address. Due to concerns with the accuracy of IP geolocation databases available online, [12] we rely on user-reported region data in this study.

4.4.3 Technical Errors

Internet measurement is susceptible to a variety of technical errors. Factors such as unreliable connectivity and poor quality of service led to an error rate of about 5.2% in our 8.9 million measurements as detailed in 1. A common way to handle such errors is to implement retries, i.e. to reattempt connections until we can be certain that the error is not caused by some transient network condition, but is actually a symptom of a website being blocked. However, in the design of our application, we found that retries caused a significant slowdown in the overall running time of our tests. We decided that a small error rate was an acceptable trade-off to make in favor of having a quicker running time, which allowed us to gain more users to contribute measurements. Having a large number of measurements from multiple vantage points on most networks allowed us to safely discard several erroneous measurements from our analysis. The error margins and how they affect the conclusions we can draw from this study are detailed in the results section below.

Error Type	Count	Description
DNS Probe Errors		
Unable to resolve host	34,356	The DNS service provider could not provide an IP address for the given host.
HTTP Probe Errors		
Timeout Exception	338,635	The website could not be reached within the configured timeout period of 2 seconds.
Connect Exception	20,930	Indicates that an error occurred while connecting to a socket.
IO Exception	2,377	Indicates that an Input/Output exception has occurred.
TLS Probe Errors		
Socket is closed	70,450	Indicates that the tests were interrupted during a run.

Table 1: Technical errors encountered.

4.4.4 Cleaning the test list

Out of a total of 10,372 potentially blocked websites, 2,087 sites were found to be unavailable, offline, or otherwise defunct. These sites were included in the DNS and TLS probes' test lists as our DNS and SNI tests do not involve the actual server associated with the hostname, and thus can detect censorship applied to offline websites as well.

As explained above, a number of technical errors arose during testing. An additional 116 websites were excluded from some of our analysis because they exhibited an error rate of greater than 85%.

Further, a small number of websites were not tested consistently across all test runs due to service outages on the unfrequented websites in our test list, and due to some users not running the complete set of tests on their devices. To minimize the impact of this limitation, 833 websites were excluded from consideration from some of our analysis.

Partial runs (wherein a device does not test all of the websites in our set) were used when analyzing overall censorship trends, such as censorship techniques, but were excluded from our analysis when comparing censorship across ASes and regions, to allow for accurate comparisons.

Overall, the full test list of 10,372 websites was used to analyze broader trends in censorship, and a shorter list of 7336 websites was used to conduct comparative analysis. The

effects of this on the conclusions we can draw from this study are detailed in the results section below.

4.5 Limitations

Our methodology suffers from the following limitations.

IP/TCP-based blocking: One limitation of our work is that we do not test for blocking of connections based on the IP address. The rationale for this decision while beginning this study was that previous work had found no evidence of IP-based blocking in India. [22, 26] However, more recent news reports and leaked government orders revealed at least two instances of the government asking ISPs to block specific IP addresses. IP-based blocking seems to be a rare scenario in India, which unfortunately our work does capture. Similarly, our work fails to account for the possibility that TLS- and HTTP-based filtering may also be conditional on the IP address.

Specific page blocking: State authorities may also issue orders for blocking of specific pages on a website. In our analysis, we only consider hostname-level analysis, and do not consider or analyze blocking of specific webpages.

ServerHello-based blocking for TLS connections: A limitation of our SNI test is that we rely solely on connecting to a known server which behaves in expected ways. In TLS 1.2 and prior versions, the server certificate travels over the wire in plaintext and can be used by the ISP to identify the website a user is trying to connect to (instead of or in addition to the SNI). While no previous work has reported that ISPs in India are the ServerHello information to censor TLS flows, our test does not capture this theoretical possibility.

NXDomain-based blocking for DNS: Due to a technical issue in our app, we were unable to distinguish between an `NXDomain` DNS error, which is indicative of censorship[22], and generic DNS errors. As detailed in 1, we encountered a very small number (34,356) of DNS errors, some of which were likely instances of censorship that we were unable to categorise.

5 Results: Overview

5.1 Censorship techniques

We observed HTTP-based blocking in 64 out of 71 ASes that we gathered measurements from, making it the most popular censorship technique amongst Indian ISPs. Its use is consistent across both large and small ISPs, and across all regions.

SNI-based blocking is used by 16 ASes out of the 64 ASes that were measured. Much of the SNI blocking observed is conducted by two large ISPs – Bharti Airtel and Reliance Jio. Two smaller ISPs, Hathway and Den Digital, were found making limited use of SNI-detection to block HTTPS connections. Ten other ISPs exhibited a very small number of SNI blocks, which is likely a result of collateral censorship. Table 2 shows which ASes conduct SNI blocking, excluding 10 ASes which showed fewer than 100 confirmed blocks.

Autonomous System (AS)	Measurements	Confirmed Blocks
AS55836 (Reliance Jio Infocomm Limited)	504,400	189,331
AS24560 (Bharti Airtel Ltd., Telemedia Services)	540,425	158,022
AS45609 (Bharti Airtel Ltd. AS for GPRS Service)	442,775	153,619
AS17488 (Hathway IP Over Cable Internet)	121,750	19,706
AS45184 (Den Digital Entertainment Pvt. Ltd. AS ISP india)	11,400	3,582
AS9829 (National Internet Backbone)	188,250	228

Table 2: Networks conducting SNI Blocking.

DNS-based blocking also finds use by larger ISPs. Only 10 out of the 64 measured ASes displayed signs of DNS blocking, with two large ASes, Atria Convergence Technologies and National Internet Backbone (the AS for the state-owned BSNL telecom provider) showing significant use of DNS blocking as shown in Table 3.

8.45% of roughly 3 million DNS measurements were conducted using known public DNS resolvers (belonging to Cloudflare, Google, Quad9, Cisco OpenDNS). Out of the 15 ASes tested using public DNS resolvers, Atria Convergence Technologies ('AS18209' and 'AS24309') was found conducting DNS injection attacks on public DNS resolvers to block websites.

ASes were also found deploying multiple blocking methods - all ASes that used SNI and DNS-based blocking also used HTTP-based blocking.

The variation in censorship methods highlighted above are important because they affect how end-users experience censorship. HTTP-based blocking, which is the most common method seen across almost all ASes measured, is ineffective. Over the last decade, more than 90% of the connections on the web have moved to encrypted communication channels, [7] with all new versions of web browsers using the encrypted HTTPS protocol by default. This tells us that most blocking on these networks is easily bypassed, or that they rely on IP-based or ServerHello-based blocking methods (detailed in section 4.5), which are largely undocumented in the Indian context and could not be detected by our network tests.

The adoption of DNS and SNI-based blocking by larger ISPs can be explained by the fact that it is more difficult to circumvent for laypeople. Note that SNI-based blocking does not allow the presentation of a censorship notice, and even with DNS-based blocking, a notice will not appear unless a website is loaded over the non-default HTTP protocol.

Like previous work, [22] we also encountered several cases

Autonomous System (AS)	Measurements	Confirmed Blocks
AS24309 (Atria Convergence Technologies Pvt. Ltd.)	289,154	125,154
AS9829 (National Internet Backbone)	176,925	92,653
AS18209 (Atria Convergence Technologies pvt ltd)	37,275	17,404
AS55577 (Atria Convergence Technologies pvt ltd)	34,800	12,408
AS131269 (ACTFIBERNET Pvt Ltd)	10,849	4,428
AS45820 (Tata Teleservices ISP AS)	11,450	4,215
AS18207 (YOU Broadband & Cable India Ltd.)	22,450	1,115
AS45609 (Bharti Airtel Ltd. AS for GPRS Service)	430,111	241
AS55836 (Reliance Jio Infocomm Limited)	488,350	40
AS38266 (Vodafone India Ltd.)	151,990	23

Table 3: Networks conducting DNS Blocking.

of 'collateral censorship', where the blocking is conducted not by the user's ISP, but by an AS that a user's ISP diverts or peers network traffic through. For instance, we saw HTTP block pages associated with AS24560 (Bharti Airtel Ltd., Telemedia Services) on AS9829 (National Internet Backbone), and tampered DNS response servers from AS9829 (National Internet Backbone) on AS24560 (Bharti Airtel Ltd., Telemedia Services) and AS55835 (Reliance Jio Infocomm Limited).

5.2 Variation of blocklists between networks and regions

In this section, we examine the variations in blocklists between ASes and regions. If a single instance of a website is confirmed to be blocked within a particular AS or state, we mark it as censored. We also report on the number of sites with 'inconclusive' measurements and the number of 'unmeasured' sites . The inconclusive measurements stem from technical errors encountered during measurement and unmeasured sites indicate the unavailability of the websites being tested, as explained in sections 4.4.3 and 4.4.4 above.

We considered 7336 websites, out of which a total of 6787 websites were blocked by at least one AS. The table in Appendix A shows the variation in censorship between ASes.[6] We discovered significant variation in the sizes of blocklists across networks, most of which range from 5000 to 7000 websites. This indicates a non-uniform implementation of censorship orders across ISPs in the country. A few smaller ASes were found blocking fewer websites in the range of 3000 to 5000 each. Figure 1 shows the variance in blocking for four popular ASNs in our dataset.

[6] In Table 4 and Appendix A, a fractional run indicates that a device only tested some of the websites in our list.

We also encountered variation in blocklists within a single AS. This can be a sign of mis-configuration or non-uniform installation of middleboxes for censorship.

Additionally, we looked at variations in website blocking between regions in India by considering the state from which the measurement was conducted. As shown in Table 4, we did not encounter much variance in censorship between states, with most blocking between 6200 and 6600 websites. Some smaller states exhibit very little censorship, but given the small number of test runs in these regions, more data is required to draw concrete conclusions.

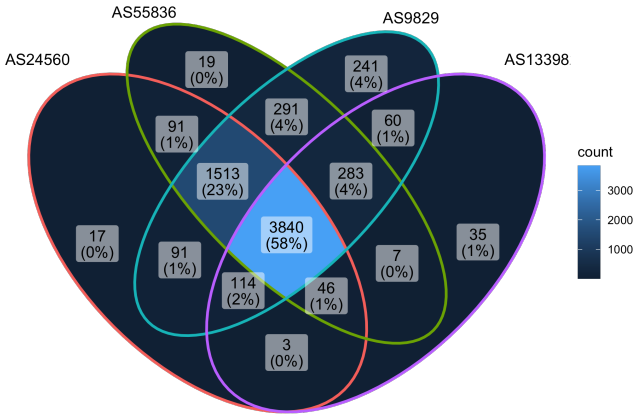

Figure 1: The intersection of websites in the blocklists for four ASNs in our data set. We notice a variation in blocking between ASNs.

6 Results: stories of non-compliance

In this section, we study specific legal orders and how various ISPs in India have complied with them. These "stories" about the implementation of online censorship document specific evidence that ISPs in India continue blocking websites and services that they are no longer legally required to. Given net neutrality regulations in India – which require ISPs to not discriminate among network traffic – these stories point to not just an unreasonable restriction of access to information online, but also violations of industry regulation.

6.1 A temporary injunction

In 2018, the producers of the Tamil-language science fiction film 2.0 approached the Madras High Court claiming that 3745 websites were hosting copyright-protected content from the film. On 27 November 2018, the court granted interim relief to the producers, and ordered ISPs in India to block those websites for a period of 16 days (from 27 November 2018

State	Runs	No. of Blocked Sites	No. of Inconclusive Sites	No. of Unmeasured Sites
Karnataka	51.7	6698	0	0
Maharashtra	51.3	6547	0	0
Delhi	36.1	6655	0	0
Kerala	22.6	6655	1	0
Tamil Nadu	19.6	6697	0	0
Haryana	19.1	6278	0	0
Telangana	16.9	6597	0	0
Uttar Pradesh	16.4	6685	0	0
West Bengal	13.9	6417	0	0
Rajasthan	13.4	6562	0	0
Punjab	7.2	6382	1	0
Assam	6.8	6390	1	0
Gujarat	6.4	6271	6	0
Uttarakhand	4.5	6440	7	0
Madhya Pradesh	4.5	6536	3	0
Chandigarh	3.4	5873	31	0
Bihar	3.3	6254	12	6
Himachal Pradesh	3.3	6396	80	0
Chhattisgarh	2.3	6217	24	19
Odisha	2.2	6259	34	1
Daman and Diu	1.1	992	919	27
Jammu and Kashmir	1.1	5958	714	10
Tripura	1.1	3366	369	39
Andhra Pradesh	0.9	5581	68	1049
Jharkhand	0.8	5448	96	1266

Table 4: Variance in website blocking between states.

to 13 December 2018). [4] The producers finally withdrew the case in early 2019. [3] In the ideal scenario then, no ISP in India should be currently blocking any of the 3745 websites. However, we found that 2370 out of 3745 were found to be blocked on at least one ISP. While it is possible that these sites were blocked through other blocking orders, this observation likely points to the fact that ISPs are not regularly updating their blocklists, and failing to comply with the precise instructions in the legal orders.

6.2 An 'unblocking'

On 18 January 2019, the Department of Telecommunications sent instructions to all Indian ISPs to unblock the website of Avaaz (`avaaz.org`), [5] ostensibly to reverse an older blocking order. As far as we know, this is the only unblocking order for which information is available outside the government and ISPs. From the data collected, we see that we have network tests from 63 ASNs that check whether `avaaz.org` was blocked on the network. We find that three ASNs still continue to block the website. These are Hathway IP Over Cable Internet (AS17488), YOU Broadband Cable India Ltd. (AS18207) and RailTel Corporation of India Ltd (AS24186).

6.3 Arbitrary blocking

In many instances, only a minority of the ISPs are blocking a website or service. Given the diversity of ASes in our dataset (71), these instances indicate either (i) unilateral action from ISPs to block a resource, which may have no legal basis; or (ii) failure to comply with 'unblocking' orders, and continuing to block websites that should now be accessible for Indian residents. The blocked websites include Youtube (`youtube.com`) which blocked on one AS; Telegram Web (`web.telegram.org`) which is blocked on 31 ASes; and Github (`github.com`) which is blocked on one AS.

7 Conclusion and Recommendations

Overall, our work presents the largest empirical study of internet censorship in India yet. Our findings also give way to policy and practical recommendations for both the government and ISPs.

Censorship techniques used by Indian ISPs: Indian law does not mandate ISPs to use any particular method of blocking, giving them some flexibility in the technical implementation of legal orders. ISPs are found to be using many censorship methods that are not transparent to users. For instance, the use of SNI-based censorship techniques means that the ISP is not able to present a censorship notice to users. Even when it comes to DNS and HTTP-based censorship, we found many ISPs are not informing users that the website they are trying to access is censored – even in cases where they have the technical ability to do so. Given the flexibility the law provides to ISPs in being free to choose the technical method of blocking, ISPs should advance transparency of censorship to the general population, so that people can challenge unjust censorship in courts.

Variances in blocklists across ISPs: We find that different ISPs are blocking different websites in India, despite ostensibly being served the same legal orders. The differences in blocklists (out of the roughly 8000 we analyzed) were as large as 2000. This affirms the findings of previous work that found that all ISPs are not blocking the same websites, [22, 8, 26] and creates abundant statistical evidence of ISPs engaging in arbitrary behavior and/or or not complying with legal orders effectively. ISPs should be more proactive in reading blocking orders in a narrow and reasonable manner, and in ensuring that temporary injunctions and unblocking orders are implemented timely.

Evidence of non-compliance with legal orders: We further contextualize blocking of specific websites with their legal orders. We find that many ISPs continue to block websites without any legal basis, violating both the rights of Indian citizens and net neutrality regulations in India. Regulators should institute a complaints mechanism, wherein users can lodge complaints against extralegal censorship by Indian ISPs.

Acknowledgments

The authors would like to thank Pooja Saxena and Akash Sheshadri for contributing to the visual design of Censorwatch; Aayush Rathi, Amber Sinha and Vipul Kharbanda for their valuable legal inputs; Internet Freedom Foundation for their support; ipinfo.io for providing free access to their data and services. The work was made possible because of research grants to the Centre for Internet and Society from the MacArthur Foundation, Article 19 and the East-West Management Institute. Gurshabad Grover's contributions were supported by a research fellowship from the Open Tech Fund.

Availability

All measurement data and the scripts used to analyse the data are available at the CensorWatch website.[7]

References

[1] Simurgh Aryan, Homa Aryan, and J. Alex Halderman. "Internet Censorship in Iran: A First Look". In: *Presented as part of the 3rd USENIX Workshop on Free and Open Communications on the Internet*. Washington, D.C.: USENIX, 2013. URL: https://www.usenix.org/conference/foci13/internet-censorship-iran-first-look.

[2] Caleb Chen. *Indian Judge clarifies that visiting blocked websites, like one for torrents, isn't illegal*. 2016. URL: https://www.privateinternetaccess.com/blog/indian-judge-clarifies-that-visiting-blocked-websites-like-one-for-torrents-isnt-illegal/.

[3] Madras High Court. *Lyca Productions Pvt. Ltd v. Bharat Sanchar Nigam Limited and Ors. (CS No. 792 of 2018)*. 2019. URL: https://www.mhc.tn.gov.in/judis/index.php/casestatus/viewpdf/438150.

[4] Madras High Court. *Order dated 27.11.2018 in Lyca Productions Pvt. Ltd v. Bharat Sanchar Nigam Limited and Ors. (CS No. 792 of 2018)*. 2018.

[5] Ministry of Communications Department of Telecommunications. *Order No. 813-7 /25/2011-DS (Vol.-VII)*. 2019.

[6] Internet Freedom Foundation. *https://internetfreedom.in/whistleblower-provides-website-blocking-orders-on-4000-websites/*. 2020. URL: https://internetfreedom.in/whistleblower-provides-website-blocking-orders-on-4000-websites/.

[7] Google. *HTTPS encryption on the web*. 2022. URL: https://transparencyreport.google.com/https/overview?hl=en.

[8] Devashish Gosain et al. "Mending Wall: On the Implementation of Censorship in India". In: *SecureComm*. Springer, 2017. URL: https://censorbib.nymity.ch/pdf/Gosain2017a.pdf.

[9] Gurshabad Grover. "To preserve freedoms online, amend the IT Act". In: *Hindustan Times* (2019). URL: https://www.hindustantimes.com/analysis/to-preserve-freedoms-online-amend-the-it-act/story-aC0jXUId4gpydJyuoBcJdI.html.

[10] Gurshabad Grover and Anna Liz Thomas. "Notes From a Foreign Field: The European Court of Human Rights on Russia's Website Blocking". In: *Indian Constitutional Law and Philosophy* (2021). https://indconlawphil.wordpress.com/2021/02/05/notes-from-a-foreign-fieldthe-european-court-of-human-rights-on-russias-website-blocking-guest-post/.

[11] Telecom Regulatory Authority of India. *Service Provider List*. 2022. URL: https://www.trai.gov.in/consumer-info/telecom/service-provider-list.

[12] Maxmind. *Geolocation Accuracy*. 2022. URL: https://support.maxmind.com/hc/en-us/articles/4407630607131-Geolocation-Accuracy.

[13] Jong Chun Park and Jedidiah R. Crandall. "Empirical Study of a National-Scale Distributed Intrusion Detection System: Backbone-Level Filtering of HTML Responses in China". In: *2010 IEEE 30th International Conference on Distributed Computing Systems*. IEEE Xplore, 2010. URL: https://ieeexplore.ieee.org/document/5541676/.

[14] Censored Planet. *Data Visualizations: Interactive visualizations by Censored Planet*. 2020. URL: https://web.archive.org/web/20200824133731/https://censoredplanet.org/data/visualizations.

[15] Reethika Ramesh et al. "Decentralized Control: A Case Study of Russia". In: *Network and Distributed System Security*. The Internet Society, 2020. URL: https://www.ndss-symposium.org/wp-content/uploads/2020/02/23098.pdf.

[16] *Rule 16, The Information Technology (Procedure and Safeguards for Blocking for Access of Information by Public) Rules*. 2009.

[17] *Section 67, The Information Technology Act*. 2000.

[18] *Section 67A, The Information Technology Act*. 2000.

[19] *Section 67B, The Information Technology Act*. 2000.

[20] *Section 69A, The Information Technology Act*. 2000.

[7] https://cis-india.github.io/censorwatch/data.html

[21] *Section 79, The Information Technology Act*. 2000.

[22] Kushagra Singh, Gurshabad Grover, and Varun Bansal. "How India Censors the Web". In: *Web Science*. ACM, 2020. URL: https://censorbib.nymity.ch/pdf/Singh2020a.pdf.

[23] Chimayi SK. *Those Unspoken Thoughts: A study of censorship and media freedom in Manipur, India*. Tech. rep. 2020. URL: https://ooni.org/documents/those-unspoken-thoughts-otf-fellow.pdf.

[24] Xueyang Xu, Z. Morley Mao, and J. Alex Halderman. "Internet Censorship in China: Where Does the Filtering Occur?" In: *Proceedings of the 12th International Conference on Passive and Active Measurement*. PAM'11. Atlanta, GA: Springer-Verlag, 2011, pp. 133–142. ISBN: 978-3-642-19259-3. URL: http://dl.acm.org/citation.cfm?id=1987510.1987524.

[25] Diwen Xue et al. "Throttling Twitter: an emerging censorship technique in Russia". In: *Internet Measurement Conference*. ACM, 2021. URL: https://dl.acm.org/doi/pdf/10.1145/3487552.3487858.

[26] Tarun Kumar Yadav et al. "Where The Light Gets In: Analyzing Web Censorship Mechanisms in India". In: *Proceedings of the Internet Measurement Conference 2018*. IMC '18. Boston, MA, USA: ACM, 2018, pp. 252–264. ISBN: 978-1-4503-5619-0. DOI: 10.1145/3278532.3278555. URL: http://doi.acm.org/10.1145/3278532.3278555.

Appendix A: Variance in website blocking amongst Autonomous Systems.

AS	AS Name	Runs	Number of Blocked Sites	Number of Inconclusive Sites	Number of Unmeasured Sites
AS24560	Bharti Airtel Ltd., Telemedia Services	51.6	5715	1	0
AS55836	Reliance Jio Infocomm Limited	48.4	6090	1	0
AS45609	Bharti Airtel Ltd. AS for GPRS Service	42.1	5747	1	0
AS24309	Atria Convergence Technologies Pvt. Ltd.	25.7	6270	0	0
AS9829	National Internet Backbone	17.3	6433	15	0
AS38266	Vodafone India Ltd.	15.5	6088	7	0
AS17488	Hathway IP Over Cable Internet	11.3	6389	1	0
AS133982	Excitel Broadband Private Limited	6.7	4388	41	0
AS23860	Alliance Broadband Services Pvt. Ltd.	6.3	4112	15	0
AS24186	RailTel Corporation of India Ltd	4.9	6228	9	0
AS45769	D-Vois Broadband Pvt Ltd	3.6	4427	42	0
AS18209	Atria Convergence Technologies pvt ltd	3.5	5777	39	0
AS133301	DWAN SUPPORTS P LTD	3.4	5250	36	0
AS55577	Atria Convergence Technologies pvt ltd	3.3	5869	29	0
AS134939	Lionet Solutions	3.3	5176	0	220
AS17465	Cable ISP in India	3.2	5128	1042	1
AS18207	YOU Broadband & Cable India Ltd.	2.4	5696	28	6
AS136696	Sifi Online Pvt Ltd	2.3	3974	214	0
AS55947	Bangalore Broadband Network Pvt Ltd	2.3	5052	71	17
AS17665	ONEOTT INTERTAINMENT LIMITED	2.3	3284	938	0
AS137098	Delix Net Solution Pvt. Ltd	2.2	3811	442	4
AS132934	Skymax broadband services Pvt. Ltd	2.2	3884	221	19
AS133007	UCN CABLE NETWORK PVT. LTD	2.2	5550	100	0
AS58678	Intech Online Private Limited	2.1	5221	56	0
AS24554	Fivenetwork Solution India Pvt Ltd Internet	2.1	3950	250	90
AS134375	Fusionnet Web Services Private Limited	2.0	5094	235	0
AS134540	Tata Teleservices (Maharashtra) Ltd	2.0	2444	0	1049
AS136334	Vortex Netsol Private Limited	1.7	3811	138	34
AS45271	Idea Cellular Limited	1.3	5132	700	37
AS136646	Shikhar Broadband Enterprises Pvt Ltd	1.2	5028	110	106
AS136308	Deenet Services Pvt Ltd	1.2	4187	513	43
AS138754	Kerala Vision Broad Band Private Limited	1.2	4466	318	17
AS58969	Kerala Communicators Cable Limited	1.1	3112	0	824
AS140112	Commonbee Broadband Pvt Ltd	1.1	992	0	1323
AS17917	Quadrant Televentures Limited	1.1	4185	223	779
AS45184	Den Digital Entertainment Pvt. Ltd. AS ISP india	1.1	4604	941	126
AS132116	Ani Network Pvt Ltd	1.1	2133	1974	141
AS134426	Mahataa Information India Private Limited	1.1	4197	0	1090
AS134928	Spiderlink Networks Pvt Ltd	1.1	2050	1170	43
AS135815	Netrexo Communications Private Limited	1.1	2868	1568	39
AS45916	Gujarat Telelink Pvt Ltd	1.1	2530	3558	39
AS137635	J.sky Media Pvt.ltd.	1.1	4253	0	714
AS10029	SHYAM SPECTRA PVT LTD	1.1	463	1650	113
AS55352	Microscan Computers Private Limited	1.1	4312	0	785
AS137100	Netmax Broadband Services	1.1	4399	247	203
AS17747	SITI NETWORKS LIMITED	1.1	4479	395	27
AS45194	Syscon Infoway Pvt. Ltd.	1.1	3032	1334	17
AS45820	Tata Teleservices ISP AS	1.1	3652	1632	29
AS132453	TRIPLE PLAY BROADBAND PRIVATE LIMITED	1.1	4052	564	40
AS135817	Esto Broadband Private Limited	1.1	4554	156	27
AS136724	Praction Networks Pvt Ltd	1.1	3731	584	20
AS132497	DIGITAL NETWORK ASSOCIATES PRIVATE LIMITED	1.1	4934	181	39
AS45415	Vasai Cable Pvt. Ltd.	1.1	2910	1411	42
AS45528	Tikona Infinet Ltd.	1.1	3982	1360	43
AS134026	Ultranet services private limited	1.1	2531	1986	39
AS131269	ACTFIBERNET Pvt Ltd	1.0	4669	0	867
AS135685	INET FIBER PVT LTD	1.0	4618	163	462