Online censorship in India, Pakistan and Indonesia

Gurshabad Grover
17 January 2024

Background

Section 69A, 79

in the IT Act

Governments, courts can pass orders

to ISPs to block websites

Central Government Act

Section 69A in The Information Technology Act, 2000

83 [69A Power to issue directions for blocking for public access of any information through any computer resource. -

https://indiankanoon.org/doc/10190353/

Central Government Act

The Information Technology (Procedure and Safeguards for Blocking for Access of Information by Public) Rules, 2009.

THE INFORMATION TECHNOLOGY (PROCEDURE AND SAFEGUARDS FOR BLOCKING FOR ACCESS OF INFORMATION BY PUBLIC) RULES, 20091

16 Requests and complaints to be confidential. Strict confidentiality shall be maintained regarding all the requests and complaints received and actions taken thereof.

https://indiankanoon.org/doc/136292737/

Central Government Act

Section 79 in The Information Technology Act, 2000

95 [79 Exemption from liability of intermediary in certain cases. -

(b) upon receiving actual knowledge, or on being notified by the appropriate Government or its agency that any information, data or communication link residing in or connected to a computer resource, controlled by the intermediary is being used to commit the unlawful act, the intermediary fails to expeditiously remove or disable access to that material on that resource without vitiating the evidence in any manner.

https://indiankanoon.org/doc/844026/

Background

Ethical and legal consideration:

Indian law doesn’t prohibit accessing blocked websites

Background

Ethical and legal consideration:

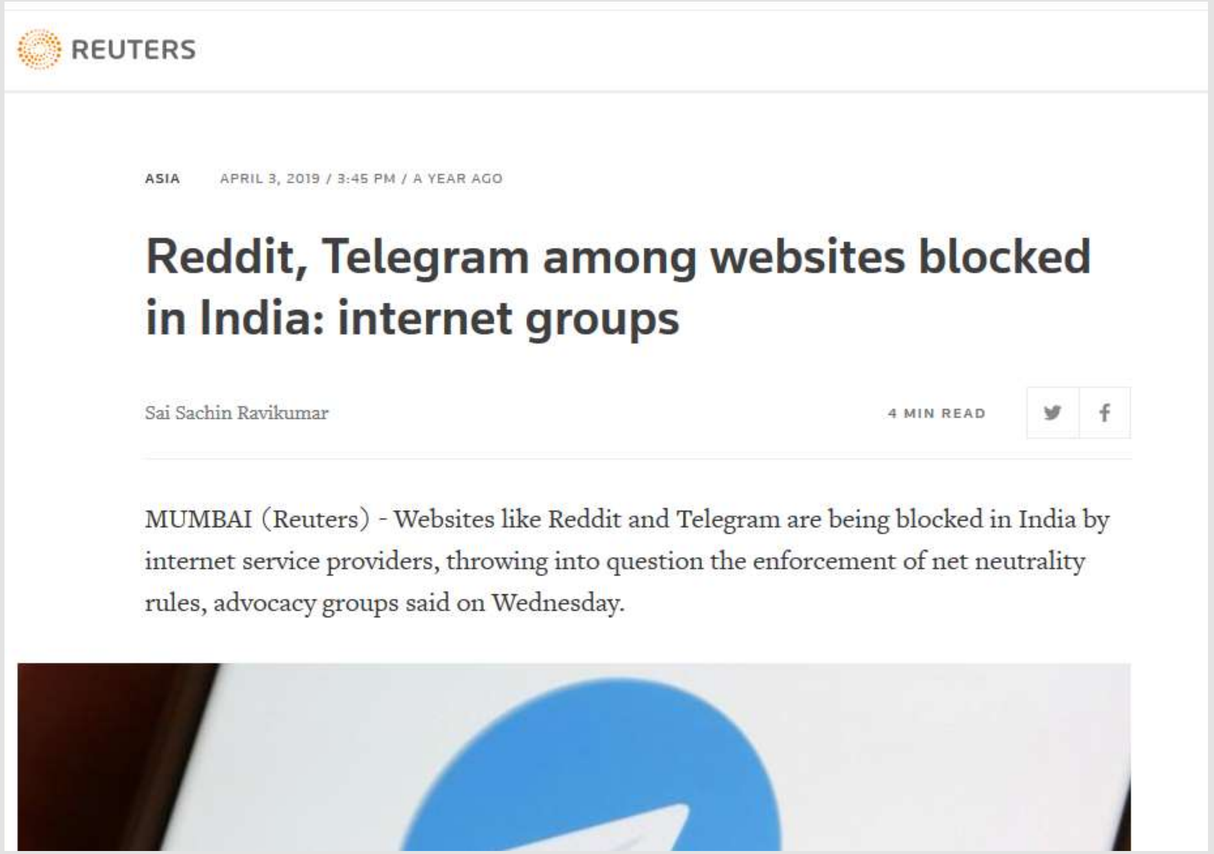
Indian law doesn’t prohibit accessing blocked websites

Censorship notices

... sometimes

Source: https://in.reuters.com/article/us-india-internet-idINKCN1RF14D

Reddit, Telegram blocked in April 2019?



REUTERS

ASIA APRIL 3, 2019 / 3:45 PM / A YEAR AGO

Reddit, Telegram among websites blocked in India: internet groups

Sai Sachin Ravikumar 4 MIN READ

MUMBAI (Reuters) - Websites like Reddit and Telegram are being blocked in India by internet service providers, throwing into question the enforcement of net neutrality rules, advocacy groups said on Wednesday.

Source: https://in.reuters.com/article/us-india-internet-idINKCN1RF14D

Reddit, Telegram blocked in April 2019?

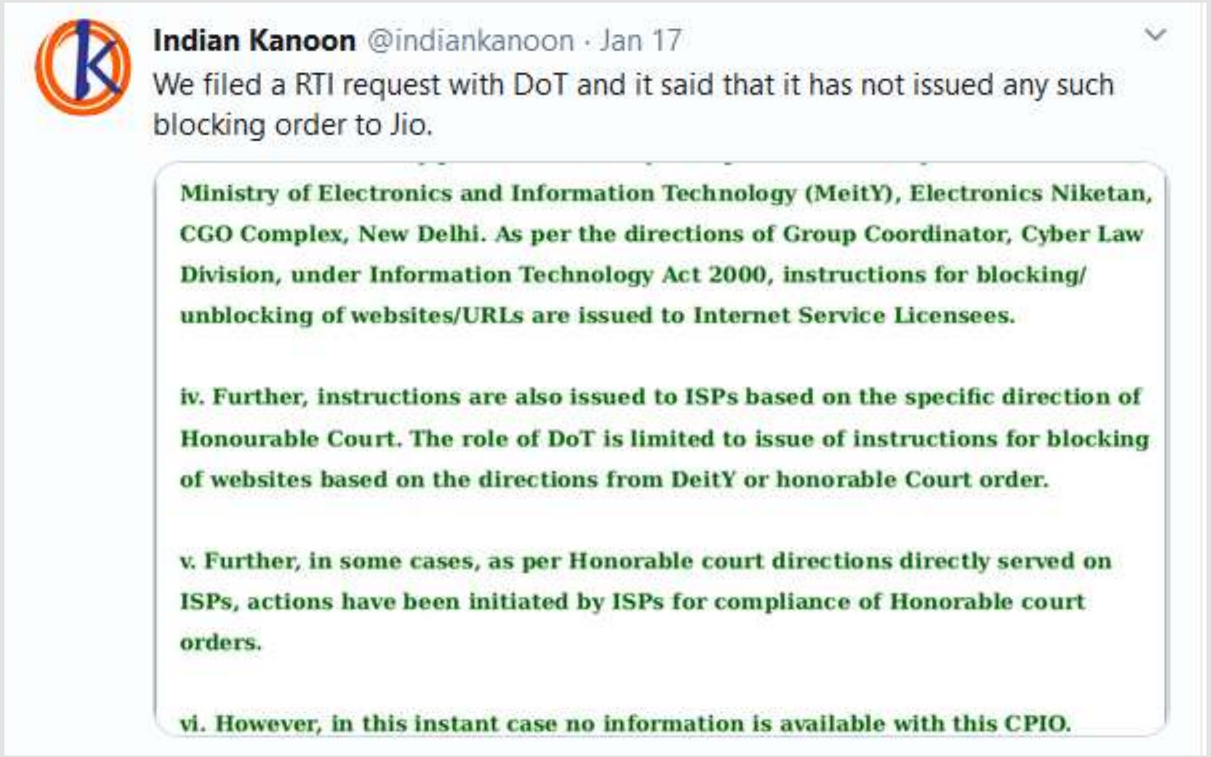
REUTERS

After complaints from Jio's internet users, Indian Kanoon founder Sushant Sharma said he had been told by Jio the portal was blocked for one day last week due to a government order.

"By evening, apparently, that order was taken back," said Sharma, whose website has some 150,000 daily visitors.

Source: https://in.reuters.com/article/us-india-internet-idINKCN1RF14D

The curious blocking of IndianKanoon.org



Source: https://twitter.com/indiankanoon/status/1218193372210323456

dowrycalculator.com

Research questions

1. What methods are ISPs using to block websites?

2. Are all ISPs blocking the same websites?

~~telegram.org~~ ~~indiankanoon.org~~

~~collegehumor.com~~ ~~thepiratebay.org~~

~~reddit.com~~ ~~dowrycalculator.com~~

Related work

- Related studies done for China, Pakistan, Syria, Italy, Iran and Korea

- Monitoring tools: OONI, Censored Planet, Censmon

Motivation

- Most work on web censorship work has focused on documenting centralized mechanisms (Iran, China)

- Very few studies on decentralised mechanisms (Pakistan, and recently Russia)

- Only one earlier study in India: Yadav, et al “Where The Light Gets In: Analyzing Web Censorship Mechanisms in India” in 2018

- No large scale study on inconsistency in website blocklists across ISPs

Methodology: data collection

Creating a list of potentially blocked websites

1 **Publicly-available or leaked government orders**

2 **Court orders**

3 **User reports***

Source: https://bit.ly/iffuserreport

Methodology: data collection

Creating a list of potentially blocked websites

1	Publicly-available or leaked government orders		890 URLs
2	Court orders		9367 URLs
3	User reports*		62 URLs

Source: https://bit.ly/iffuserreport

Methodology: data curation

Creating a list of potentially blocked websites

1	Publicly-available or leaked government orders		890 URLs
2	Court orders		9367 URLs
3	User reports*		62 URLs

9673 URLs

(after removing duplicates)

Methodology: data curation

4379
Blocked websites

This is the largest known corpus of potentially blocked hostnames in India.

Methodology: ISPs

Six major ISPs in India

ACT | Airtel | BSNL | Jio

MTNL | Vodafone

Methodology: data curation

98.82%
of internet subscribers
in India

The Telecom Regulatory Authority of India reveals that as of October 2019, these six ISPs together serve 657.46 million users.

Methodology: DNS

DNS (uncensored)

Non-censorious

example.com

93.184.216.34

(Just for illustrating the correct DNS records)

Methodology: DNS

DNS (uncensored)

Non-censorious ISP

Non-censorious

example.com

93.184.216.34

(Just for illustrating the correct DNS records)

Methodology: DNS

DNS poisoning

Censorious ISP

Non-censorious

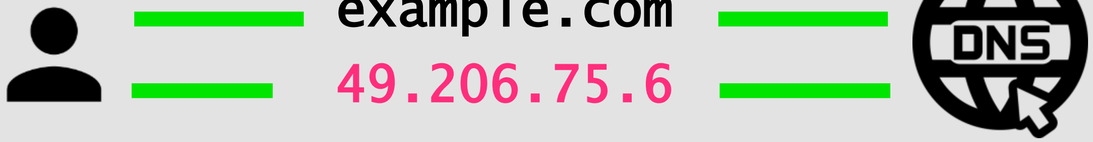
example.com

93.184.216.34

(Just for illustrating the correct DNS records)

Methodology: DNS

DNS poisoning



Censorious ISP

Non-censorious

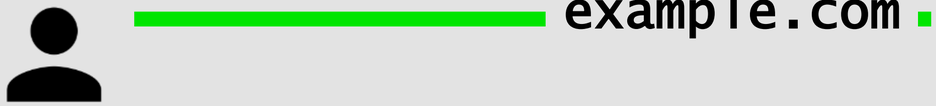
example.com

93.184.216.34

(Just for illustrating the correct DNS records)

Methodology: DNS

DNS injection



Any DNS server

example.com
93.184.216.34

Methodology: DNS

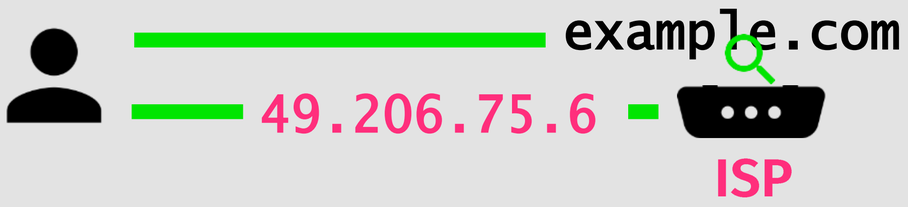
DNS injection

Any DNS server

example.com
93.184.216.34

Methodology: DNS

DNS injection



Any DNS server

example.com
93.184.216.34

Methodology: DNS (Previous work)

- Compare test resolver's response with a trusted resolver's response

  Problem: trusted resolvers can return a different IP address (legitimately)

- Lowe, et al select multiple resolvers, investigate only where response is same

  Problem: significant reduction in the size of the test list

- Yadav, et al rely on AS number

  Problems: (1) will spoofed IP address always belong to the same AS?

  (2) what if the website is hosted on the same AS?

Methodology: DNS (Proposed technique)

1. Query five trusted resolvers, and test resolver

2. If response from test resolver $\in$ {(responses from trusted resolvers)} **Not censored**

3. If response from test resolver is `NXDOMAIN` or bogon IP **Censored**

4. For others, use data from all responses: is there an IP address present with an unusually high frequency? **Censored**

   i.e. compare relative frequency of most frequent IP address

Methodology: HTTP

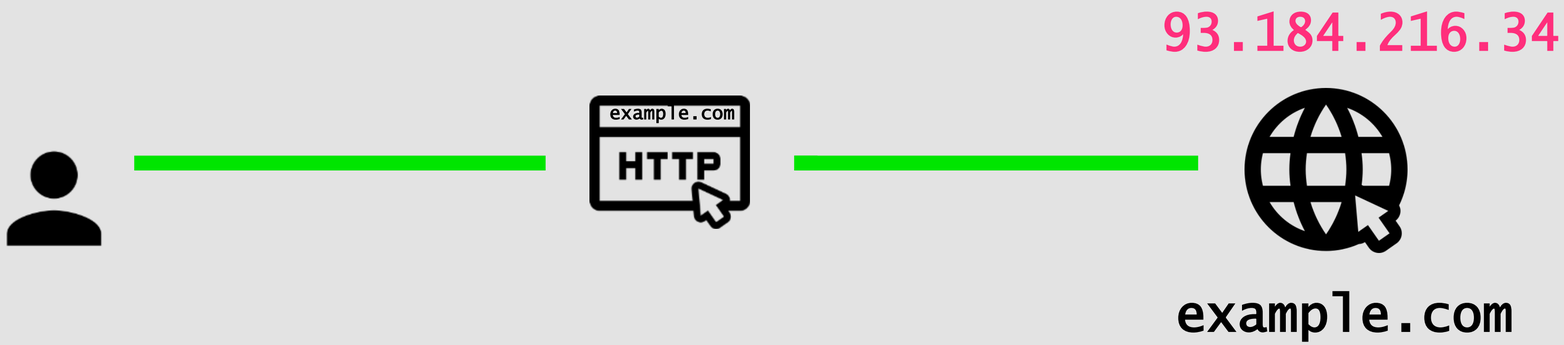
HTTP (uncensored)

93.184.216.34

example.com

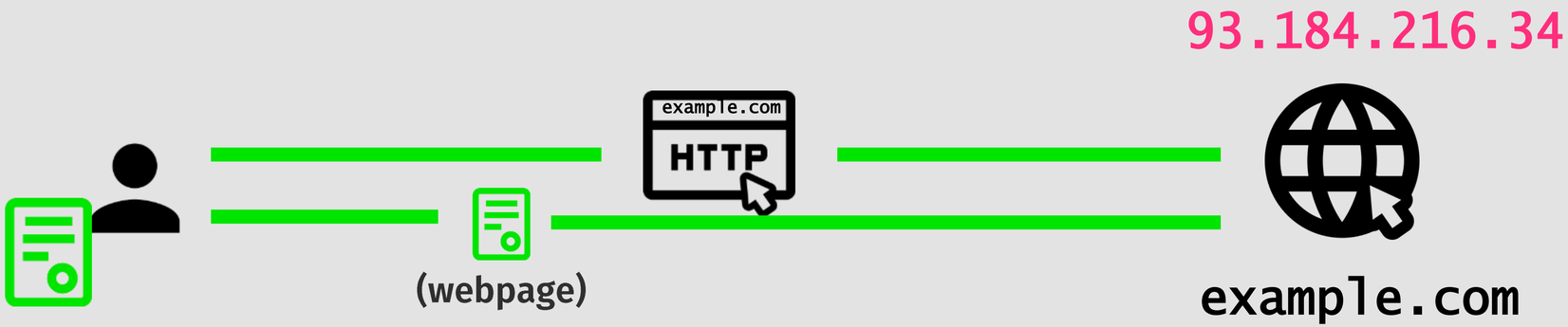
Methodology: HTTP

HTTP (uncensored)



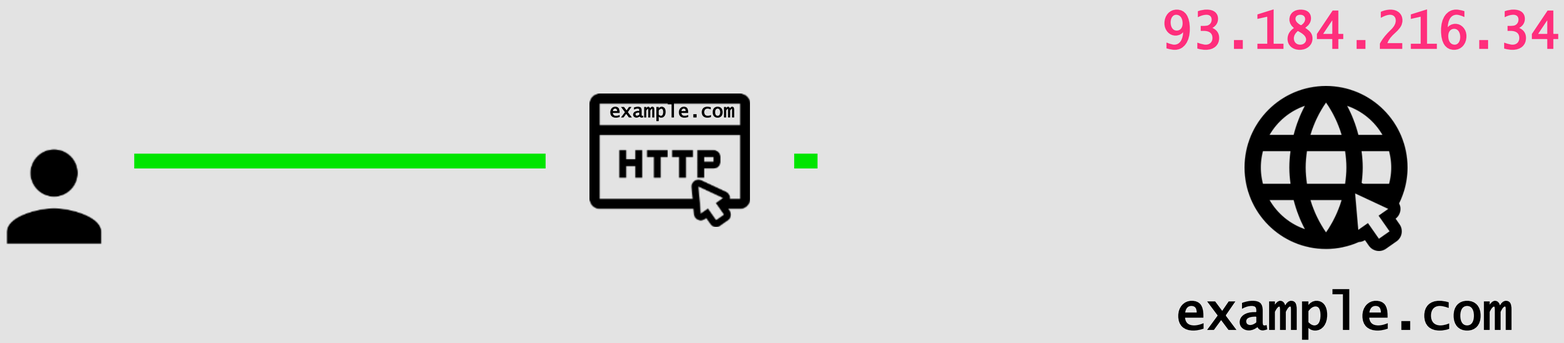
Methodology: HTTP

HTTP (uncensored)



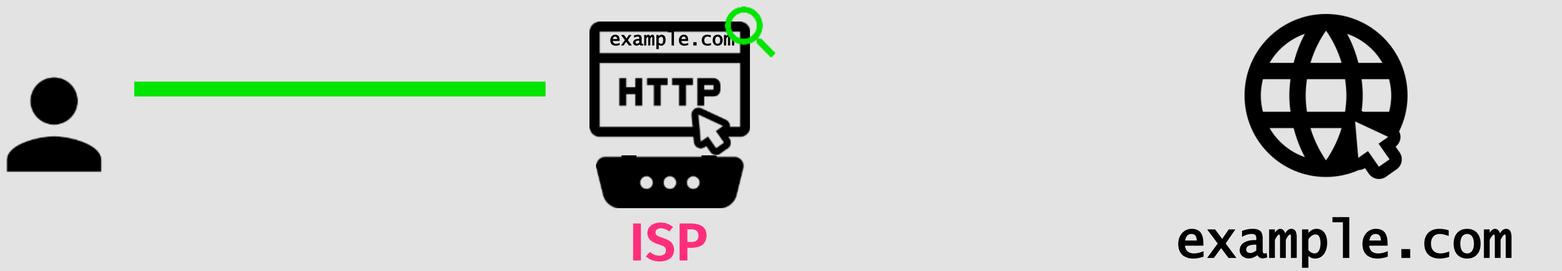
Methodology: HTTP

HTTP-based censorship



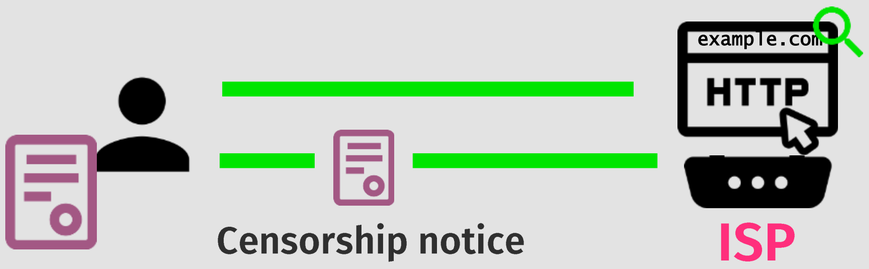
Methodology: HTTP

HTTP-based censorship



Methodology: HTTP

HTTP-based censorship



93.184.216.34

example.com

Methodology: HTTP (Previous work)

- Simple comparison of responses with uncensored responses collected via controls

  Problem: Content often keeps changing, content may be localised

- Jones, et al rely on length and structure of responses to detect censorship notices

  Problem: Assumption of censorship notices

- OONI does a more elaborate comparison (status codes, headers, lengths)

  Problems: Not a lot, but Yadav et al found lots of false negatives and positives for India

Methodology: HTTP (Proposed technique)

1. Resolve hostname and get a response via test and 5 control networks

2. If status codes (Success, Redirection, Error) do not match — Censored
   (vice versa may not be true though)

3. If Success (2xx), and response length, bodies do not match — Censored
4. If Redirection (3xx), and domain name in redirect URL do not match — Censored
5. If Error (4xx or 5xx), and session header keys do not match — Censored

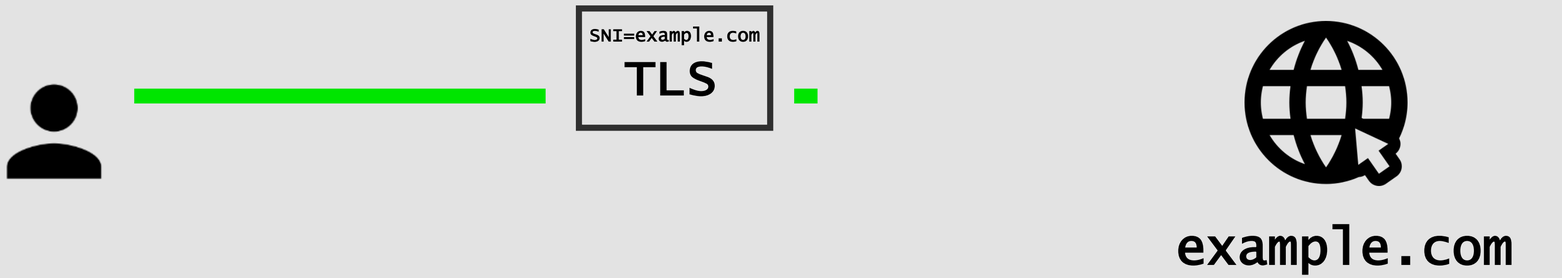
Methodology: HTTP (Proposed technique)

To verify our method's accuracy, we manually inspected and compared against 500 responses

Detection Technique	Precision		Recall		F1 score	
	C	U	C	U	C	U
Length difference [28, 47]	0.65	0.73	0.77	0.59	0.70	0.66
HTML similarity [28]	0.45	0.44	0.62	0.28	0.52	0.34
OONI [19]	0.67	1.00	1.00	0.54	0.80	0.70
Proposed	**0.71**	0.98	0.99	**0.63**	**0.83**	**0.77**

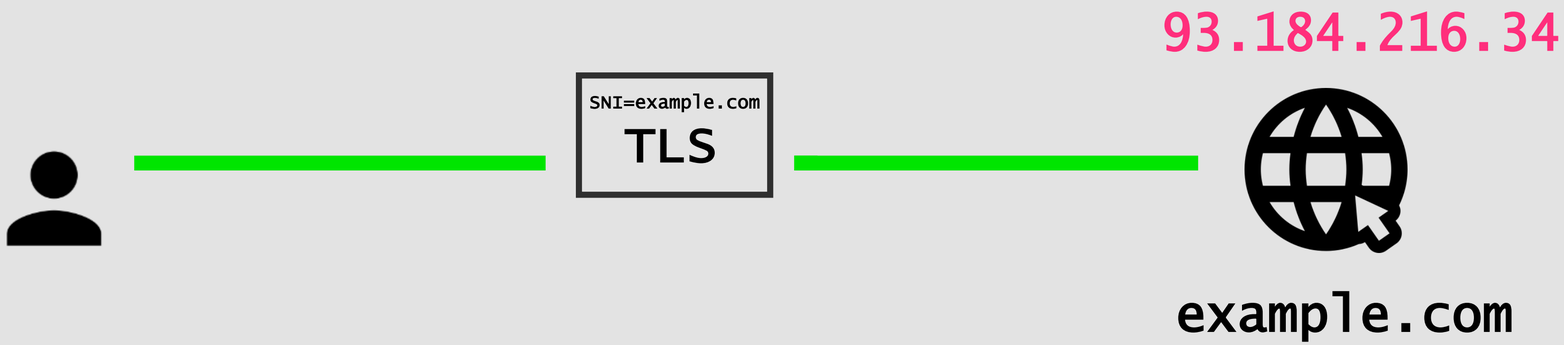
Methodology: SNI

SNI-based (uncensored)



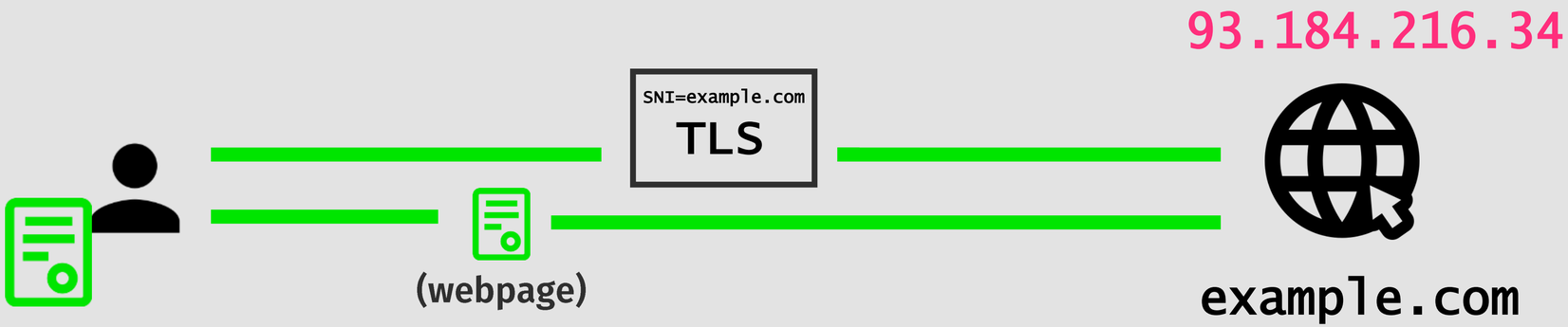
Methodology: SNI

SNI-based (uncensored)



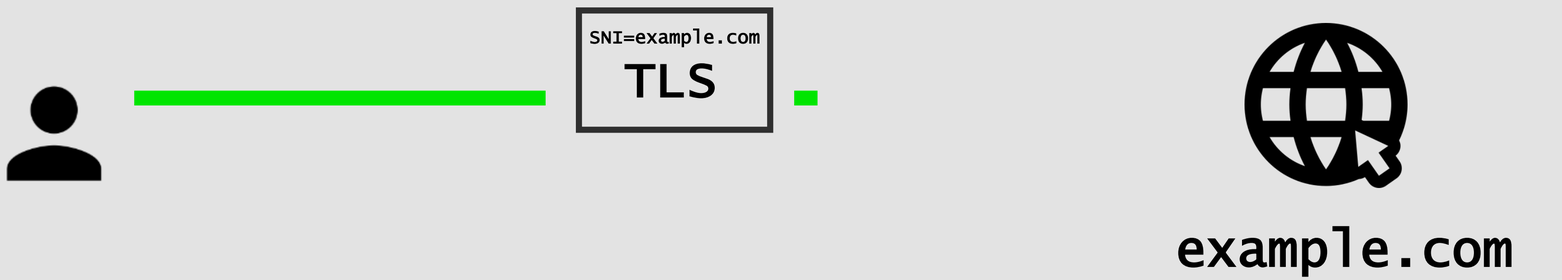
Methodology: SNI

SNI-based (uncensored)



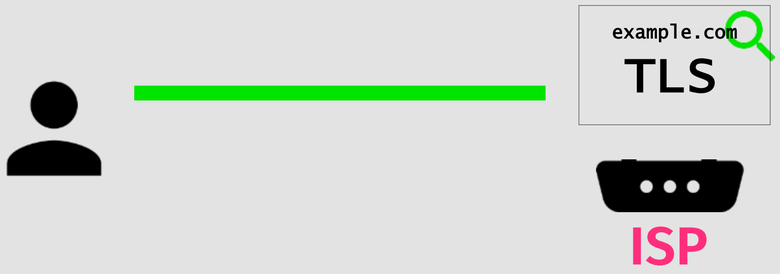
Methodology: SNI

SNI-based censorship



Methodology: SNI

SNI-based censorship

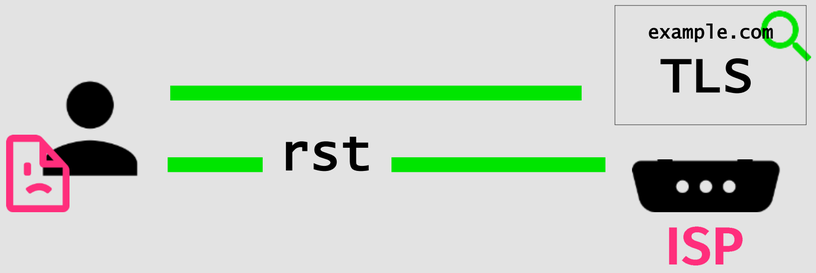


93.184.216.34

example.com

Methodology: SNI

SNI-based censorship



93.184.216.34

example.com

Methodology: SNI (Proposed technique)

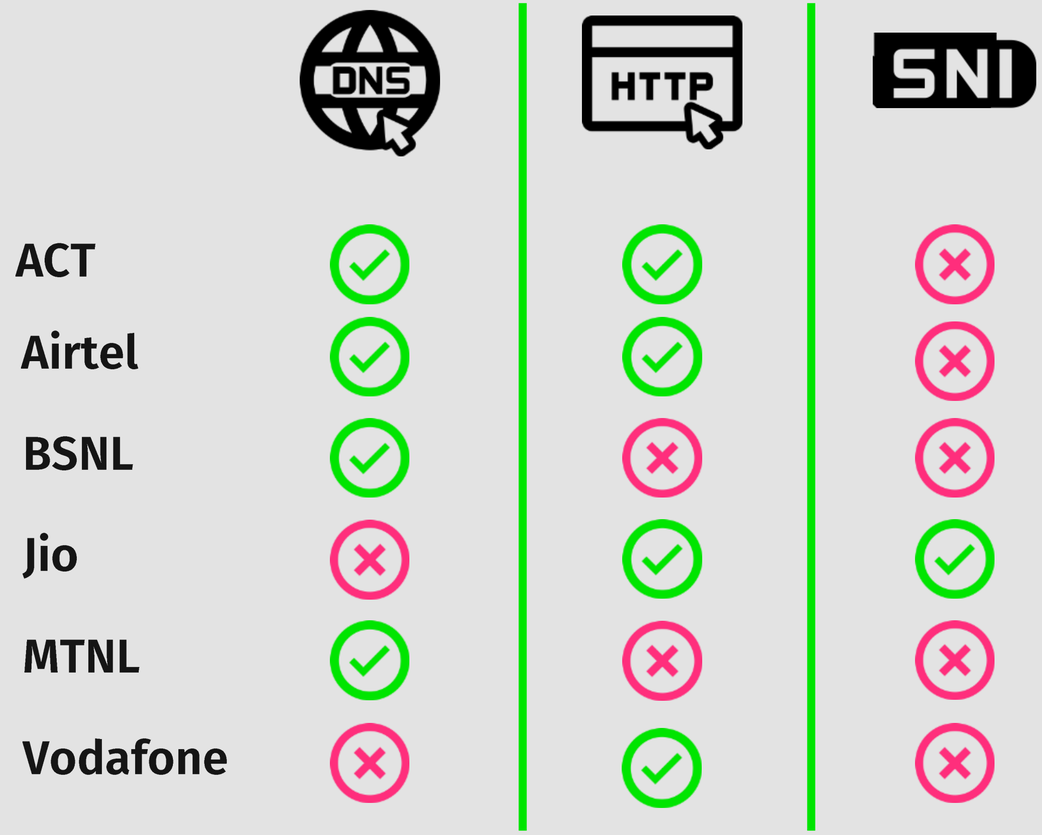
SNI

1. Set up server that accepts connections even if it doesn’t host the website present in the SNI

2. Establish TLS 1.3 connection (encrypted cert!) with our server and send SNI of potentially blocked website

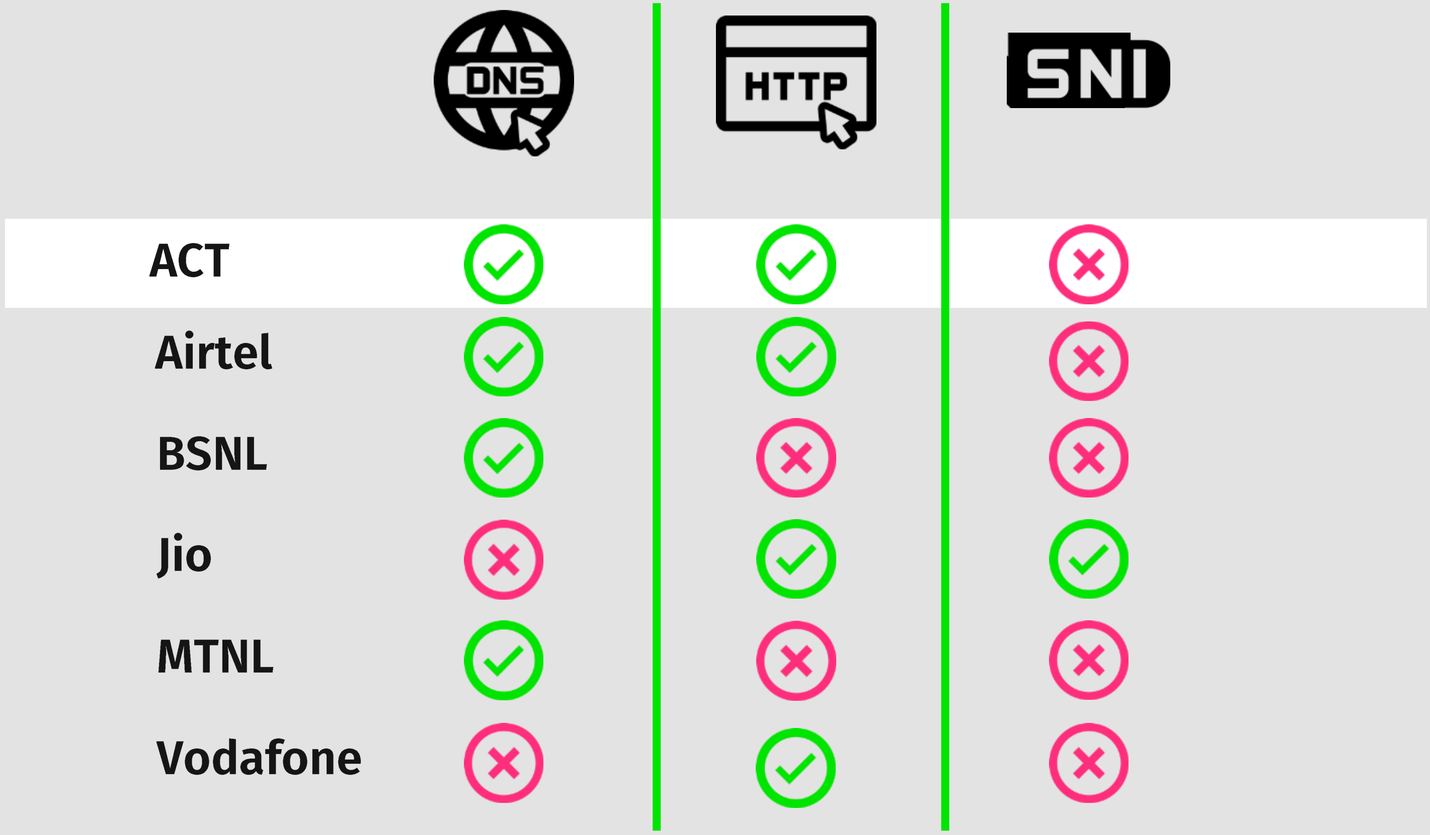
3. If you spot a failure to connect: Censored

Results: Censorship Techniques



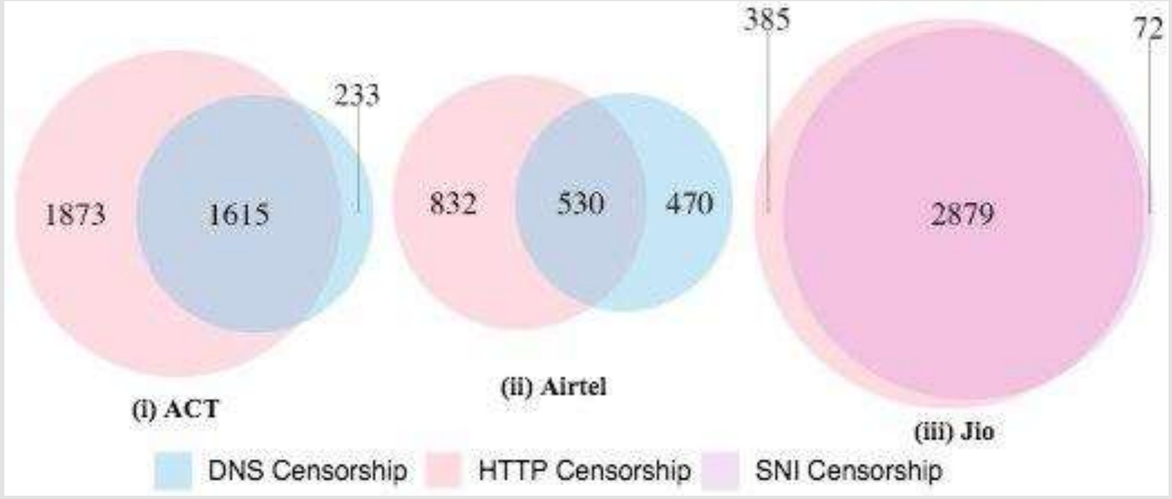
Different censorship mechanisms, individually or in combination to censor websites.

Results: Censorship Techniques



ACT: only DNS for 233, only HTTP for 1873, and both to block 1615 websites

Results: Censorship Techniques



Censorship techniques used by ACT, Airtel and Jio

Results: Censorship Techniques

- Four ISPs (ACT, Airtel, BSNL and MTNL) using DNS-based censorship
- Most are sending censorship notices, except Airtel which responds with NXDOMAIN
- No instances of collateral censorship (consistent with Yadav et al findings)

Results: Censorship Techniques

- HTTP-based censorship observed in ACT, Airtel, Jio and Vodafone
- All of them except Airtel serving censorship notices (Airtel just sends a TCP RST)
- And some collateral censorship: observed Airtel and ACT notices in BSNL and MTNL

Results: Censorship Techniques

SNI

- Results indicated that only Reliance Jio was using SNI-based blocking

- Censorship notices not possible!

Results: Website blocklists

	Websites blocked
ACT	3721
Airtel	1892
BSNL	3033
Jio	3340
MTNL	3182
Vodafone	2273

Number of websites (out of 4033) blocked by ISPs

Results: Website blocklists

just

27.64%

of all blocked websites

are blocked by all six ISPs.

1115 of 4033 websites

Results: Website blocklists

just

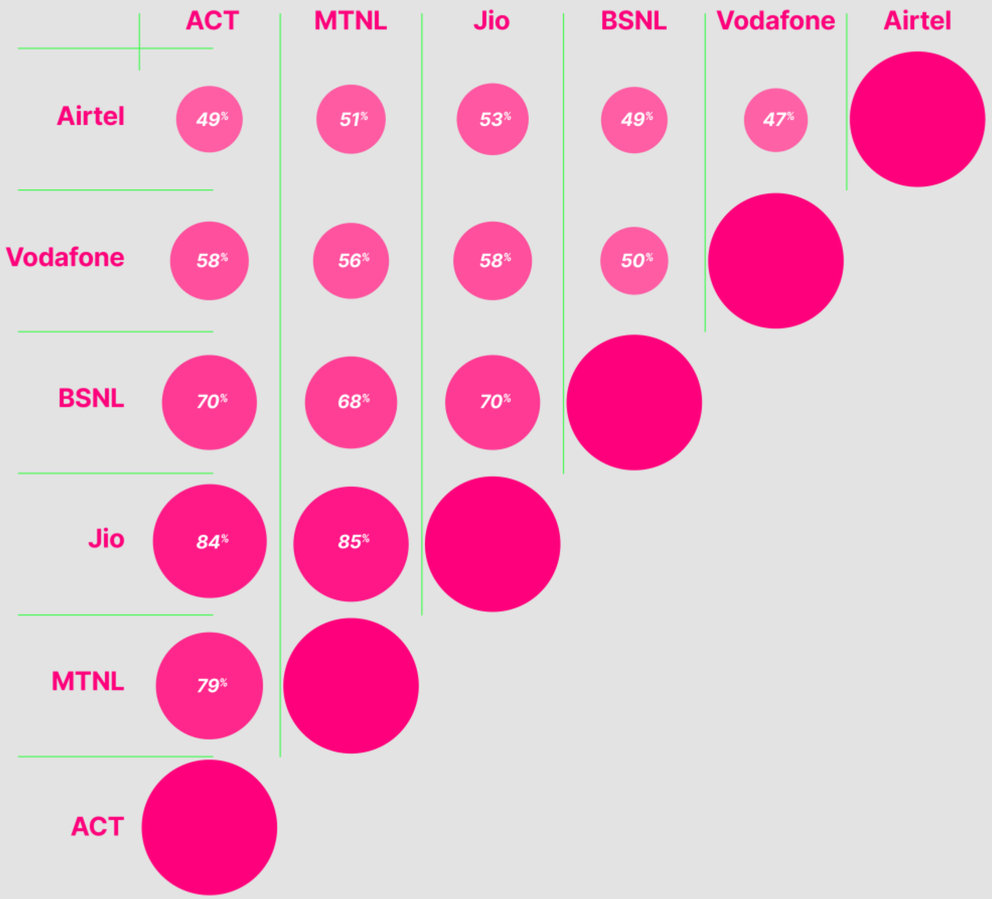
27.64%

are blocked by all six ISPs.

1115 of 4033 websites

We also found that lots of websites (215) are being blocked by only a single ISP out of the six.

Results: Website blocklists



Map illustrating the overlap of blocklists across ISPs.

For each pair of ISP blocklists L_a and L_b

$$\frac{|L_a \cap L_b|}{|L_a \cup L_b|}$$

Results: Website blocklists

ISPs are either

1 Not properly complying with website blocking (or subsequent unblocking orders).

and/ or

2 Arbitrarily blocking websites without the backing of a legal order.

Results: Website blocklists

ISPs are either

1 Not properly complying with website blocking (or subsequent unblocking orders).

and/ or

2 Arbitrarily blocking websites without the backing of a legal order.

India's net neutrality regulations prohibit such behaviour

Conclusion

1. Need to re-evaluate legal and technical mechanisms of web censorship in India

Conclusion

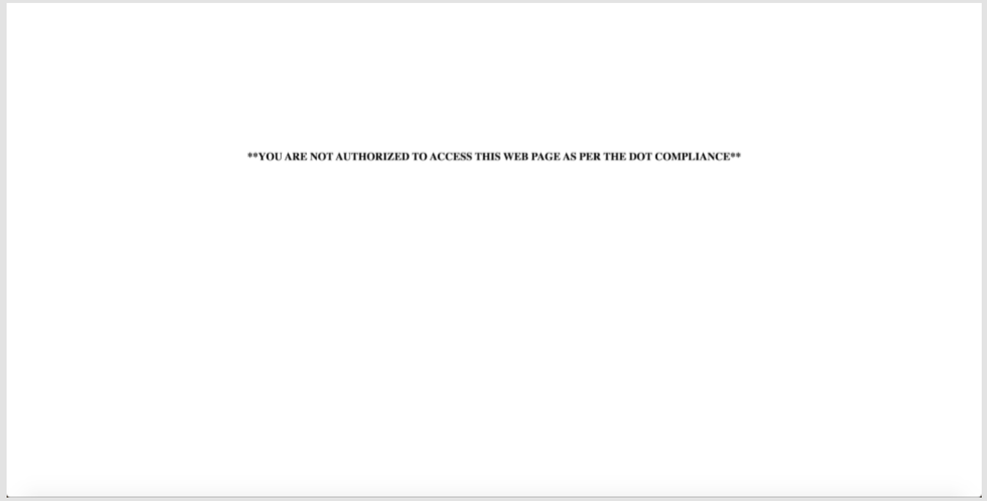
1. Need to re-evaluate legal and technical mechanisms of web censorship in India

2. Have a net neutrality monitoring mechanism in place

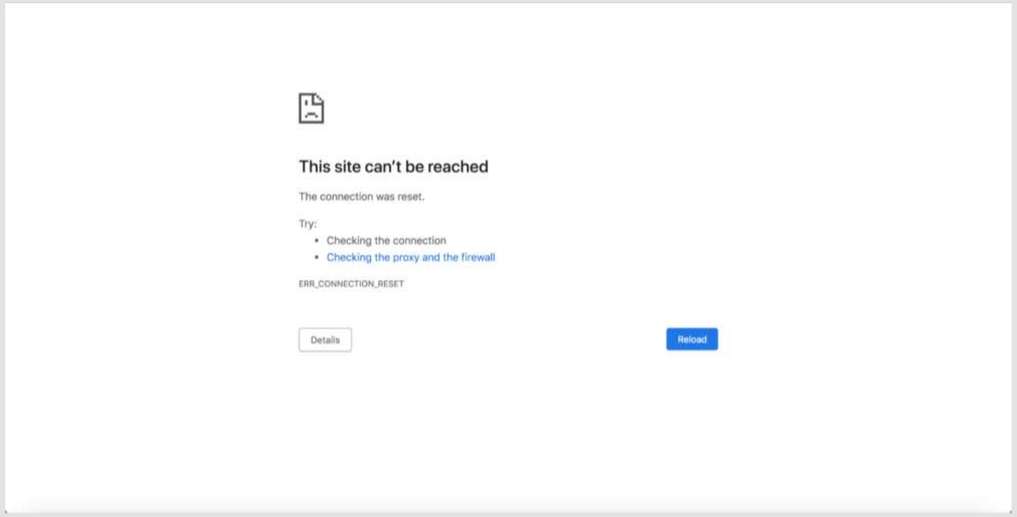
Conclusion

1. Need to re-evaluate legal and technical mechanisms of web censorship in India
2. Have a net neutrality monitoring mechanism in place
3. ISPs should use transparent blocking methods

Censorship notices



HTTP-based blocking on Jio



SNI-based blocking on Jio

Conclusion

1. Need to re-evaluate legal and technical mechanisms of web censorship in India
2. Have a net neutrality monitoring mechanism in place
3. ISPs should use transparent blocking methods

Censorwatch

- Compiled list of potentially blocked websites
  - Court orders
  - Leaked and public government orders
  - User reports

- Developed mobile application: DNS, HTTP and TLS tests

- 10,372 websites contributed by 331 users from 25 states in India, across 71 ASes (~9 million measurements)

Censorwatch - Results

- HTTP-based blocking in 64 out of 71 ASes
- SNI-based blocking is used by 16 ASes out of the 64 ASes
- 10 out of the 64 ASes measured displayed signs of DNS blocking

Censorwatch - Results

- Significant variation in blocklists between networks, with most ranging from 5000 to 7000 websites blocked

- smaller ASes were found blocking fewer websites in the range of 3000 to 5000 each

Censorwatch - Contextualisation with legal orders

- A temporary injunction to stop piracy
- An ‘unblocking’ of awaaz.org
- Arbitrary blocking

Comparative study for India, Pakistan and Indonesia

Objectives

- Map laws, techniques and infrastructure used for online censorship in South Asia (India, Pakistan, Indonesia)

- Understand role of ISPs in exacerbating or minimising the effects of censorship (using OONI network measurement data)

India: Overview (1)

- There are both private and public internet service providers.
  - The Government of India and the courts can order internet service providers to block websites and online content.

- There are both private and public internet exchanges.
  - There is no evidence of an internet exchange being used for traffic filtering.

- The Government does not mandate any particular method of blocking, and ISPs are free to decide what method to use. There is also evidence of inconsistencies in the block-lists of different ISPs.

India: Overview (2)

- Legal background: IT Act allows blocking powers, notice sent to ISPs
- The complete official list of blocked websites is NOT available for India: all orders passed by the Government are secret because of regulations.
- Court orders are publicly available, and the Government has provided a list of such orders in response to Right to Information (RTI) requests.
- List of other potentially blocked websites have been compiled by some researchers.

Pakistan: Overview (1)

- The complete official list of blocked websites is NOT available for Pakistan.

- There are both private and public internet service providers.
  - The Government can order internet service providers to block websites and online content. ISPs are implementing filtering at this level.

- There is a state-supported/backed internet exchange. There does not seem to be any information about a private/non-state cooperative internet exchange.
  - There is evidence (from 2013) that the internet exchange may be involved in censoring traffic. But the work (from 2016) does not make reference to this mechanism.

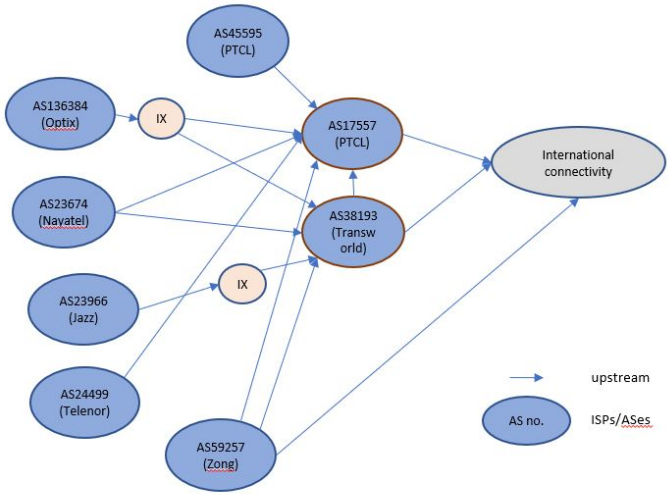
Pakistan: Overview (2)

- Centralization trend
  - Web monitoring system (tender)
  - Centralized DNS server

- Web Monitoring System: government acknowledged operationalisation in 2020

- Lines of international connectivity controlled by PTCL and Transworld

Pakistan


AS45595
(PTCL)
AS136384
(Optix)
IX
AS17557
(PTCL)
International
connectivity
AS23674
(Nayatel)
AS38193
(Transw
orld)
IX
AS23966
(Jazz)
AS24499
(Telenor)
AS59257
(Zong)
upstream
AS no.
ISPs/ASes

Indonesia: Overview (1)

- There are both private and public internet service providers.
  - Diverse interconnection and international connectivity, there does not seem to be any centralized infrastructure.

- There are examples of private internet exchanges. Currently, there is no evidence that IXPs are being used for censorship.

- Blocking orders seem to have been passed to internet service providers by multiple state authorities, including the Ministry of Communications and Information Technology (MCIT) and the Election Oversight Body.

Indonesia: Overview (2)

- There were regulations that offered technical guidance to ISPs, including following a blocklist maintained by the MCIT.

- Total number of websites blocked is available, but no details one what is blocked and why. However, the TRUST+ Positif (which ISPs were mandated to follow as per the 2014 regulations) is publicly available.

- Regulation was updated replaced in 2020: allows Indonesian ISPs to block websites on their own discretion.

Indonesia: Overview (3)

- Popular services like Vimeo and Reddit have been found to be blocked only by a chunk of Indonesian ISPs

- 2022 proposal in Indonesia Network Operators' Group (IGNOG) and Indonesia Internet Service Provider Association (APJII) presentation:
  - centralise DNS queries
  - no evidence that it has been implemented yet

Discussion (1): There are levels to it

- Censorship occurs at:
  - ISPs
  - Internet exchange points
  - Lines and chokepoints of international connectivity

- All of these can be public (state-controlled) or private infrastructure

Discussion (2): Private infrastructure

- Many ISPs are overzealous (in different ways) when it comes to blocking
- ISPs need to be pushed for adopting human rights considerations
  - (Framework)
- Measurements can uncover unlawful practices

Discussion (3): Centralized vs. decentralized

- There is no ‘great firewall’ in most cases – not even in China and Iran.
- Decentralized controls does not mean better outcomes in terms of access
- Conversely, centralized infrastructure may not necessarily mean worse outcomes

Future questions

- IETF already working in many ways (DoT, DoH, QUIC, ECH) to plug privacy-gaps

- Target of censorship orders may change
  - What happened when HTTPS was getting popular?
  - Geoblocking at open DNS resolvers?
  - Blocking of specific protocols and VPNs?