How India Censors the Web

Kushagra Singh, **Gurshabad Grover**, and Varun Bansal

(thanks to Akash Sheshadri for the slide designs!)

Human Rights Protocol Considerations Research Group @ (IETF107) Virtual Interim
08 April 2020

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

But also see: work by Censored Planet and Prof. Roya Ensafi

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?

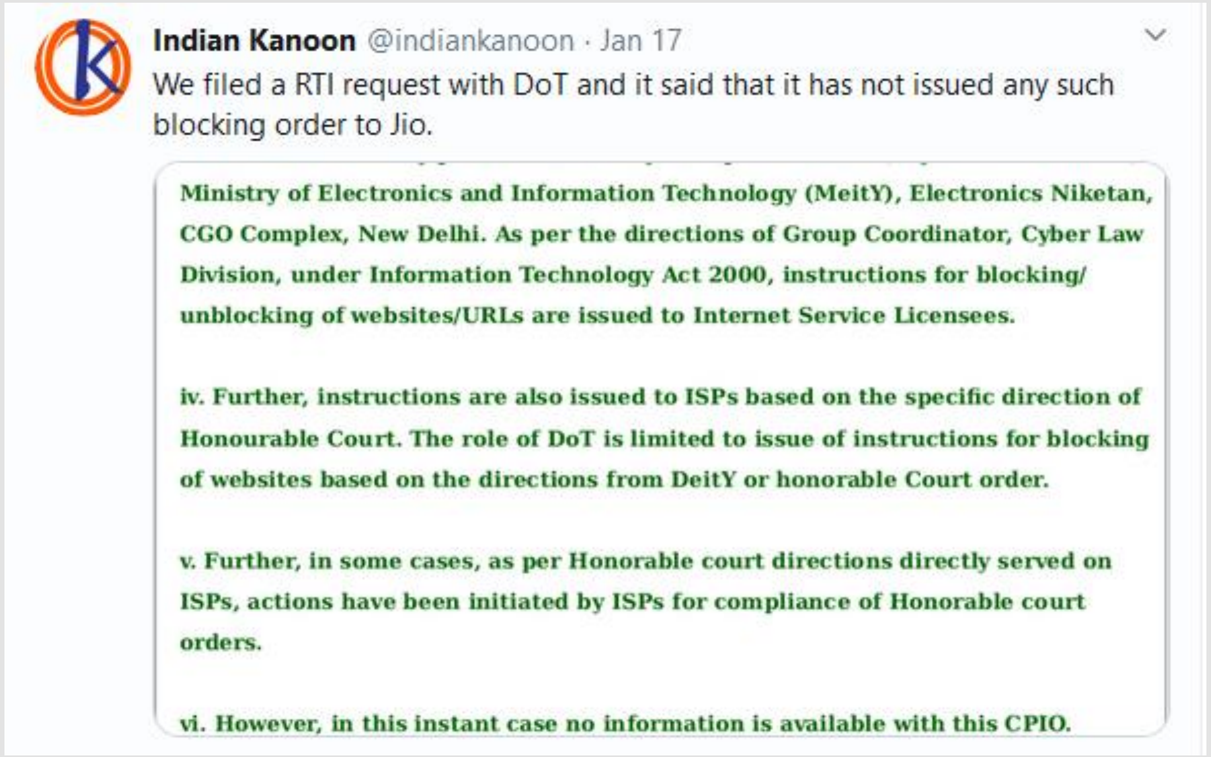
REUTERS

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.

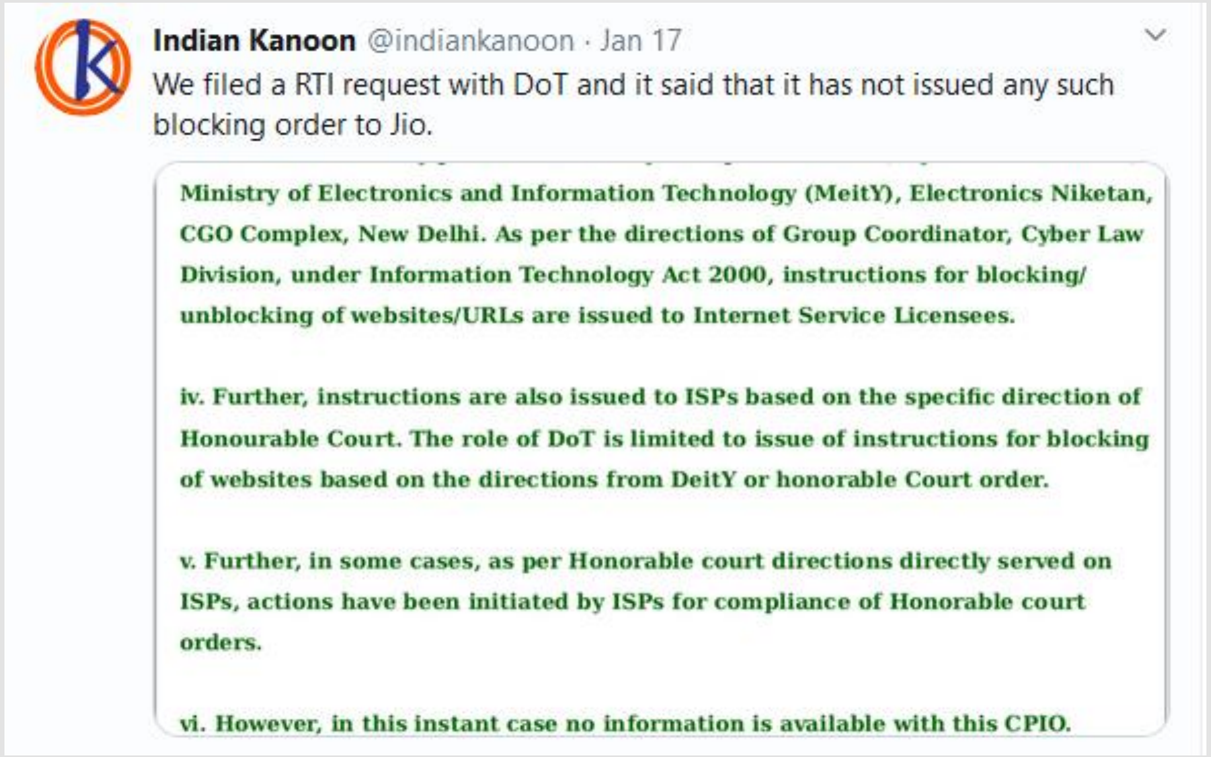
Source: https://in.reuters.com/article/us-india-internet-idINKCN1RF14D

The curious blocking of IndianKanoon.org



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

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

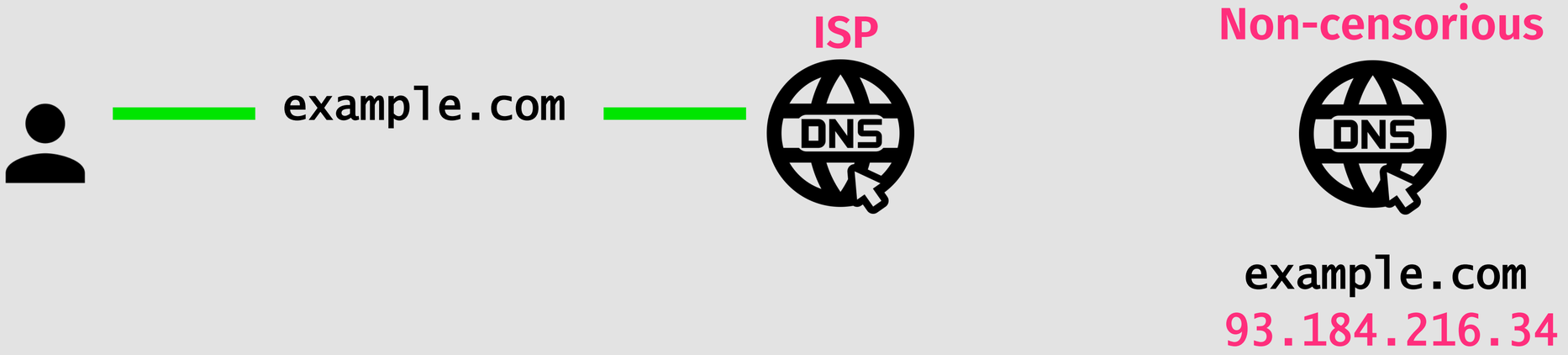
98.82%

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

1 DNS Poisoning

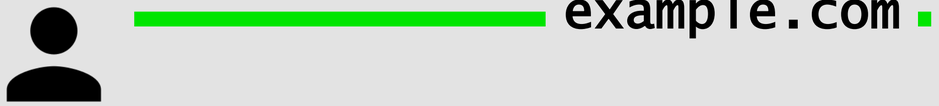


Methodology: DNS

1 DNS Poisoning

Methodology: DNS

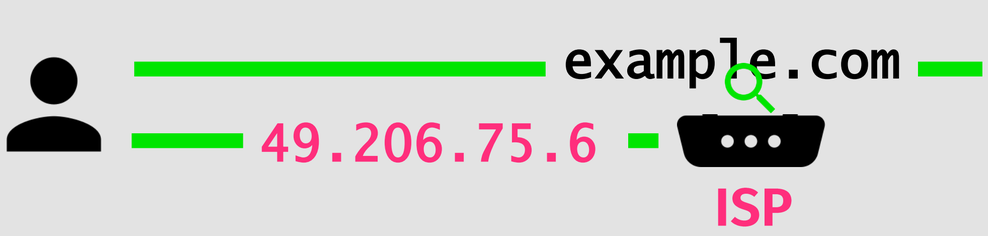
1 DNS Injection



example.com
93.184.216.34

Methodology: DNS

1 DNS Injection



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**

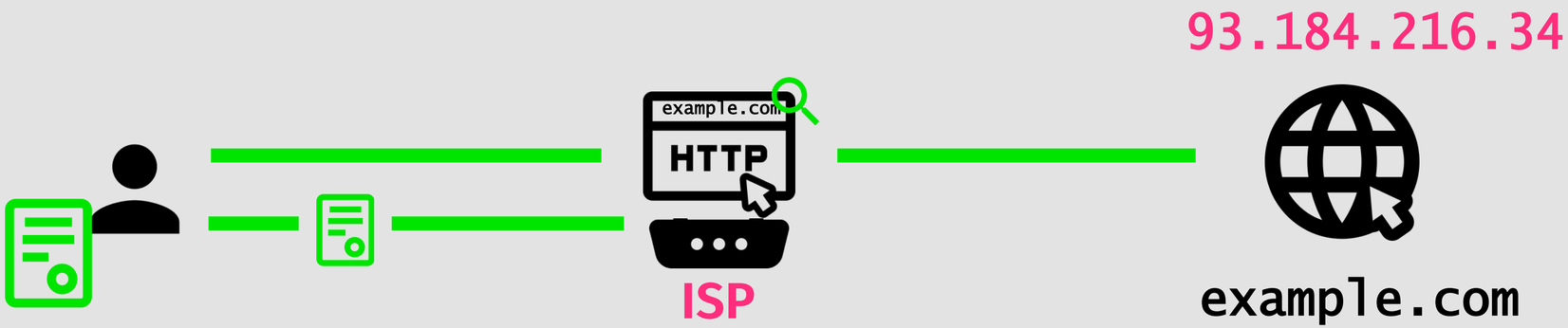
   i.e. compare relative frequency of most frequent IP address

Methodology: HTTP

93.184.216.34

example.com

Methodology: HTTP



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)

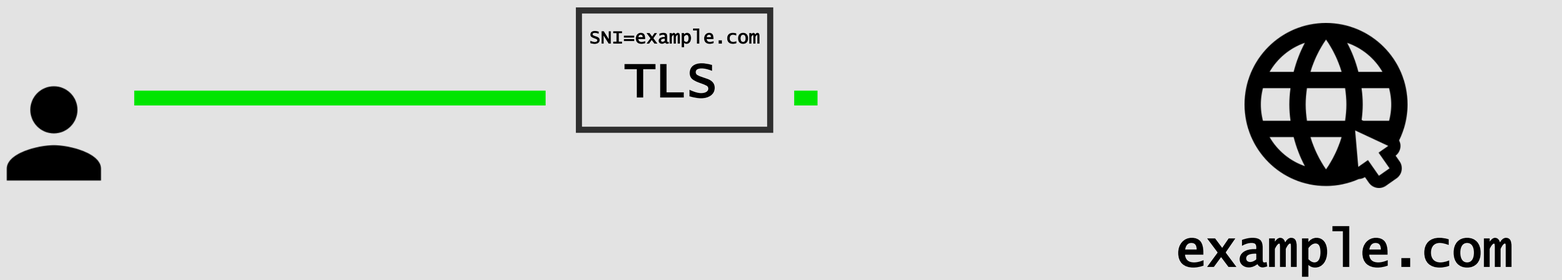
3. If Success (2xx), and response length, bodies do not match — Censored
4. If Redirection (3xx), and domain name in redirect URL do not match — Censored
5. If Error (4xx or 5xx), and session header keys do not match — Censored

Methodology: HTTP (Proposed technique)

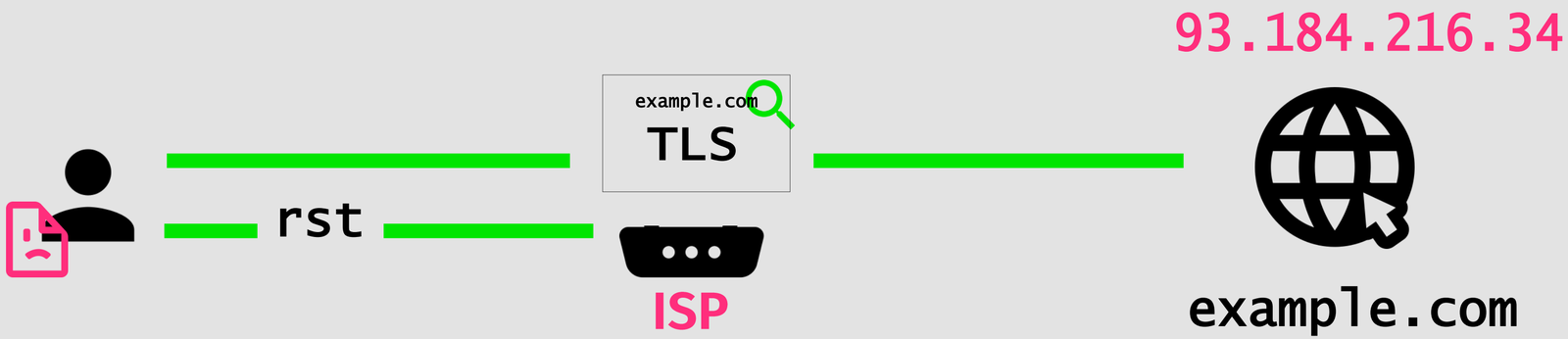
To verify our method's accuracy, we manually inspected and compared against 500 responses

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



Methodology: SNI



Methodology: SNI (Proposed technique)

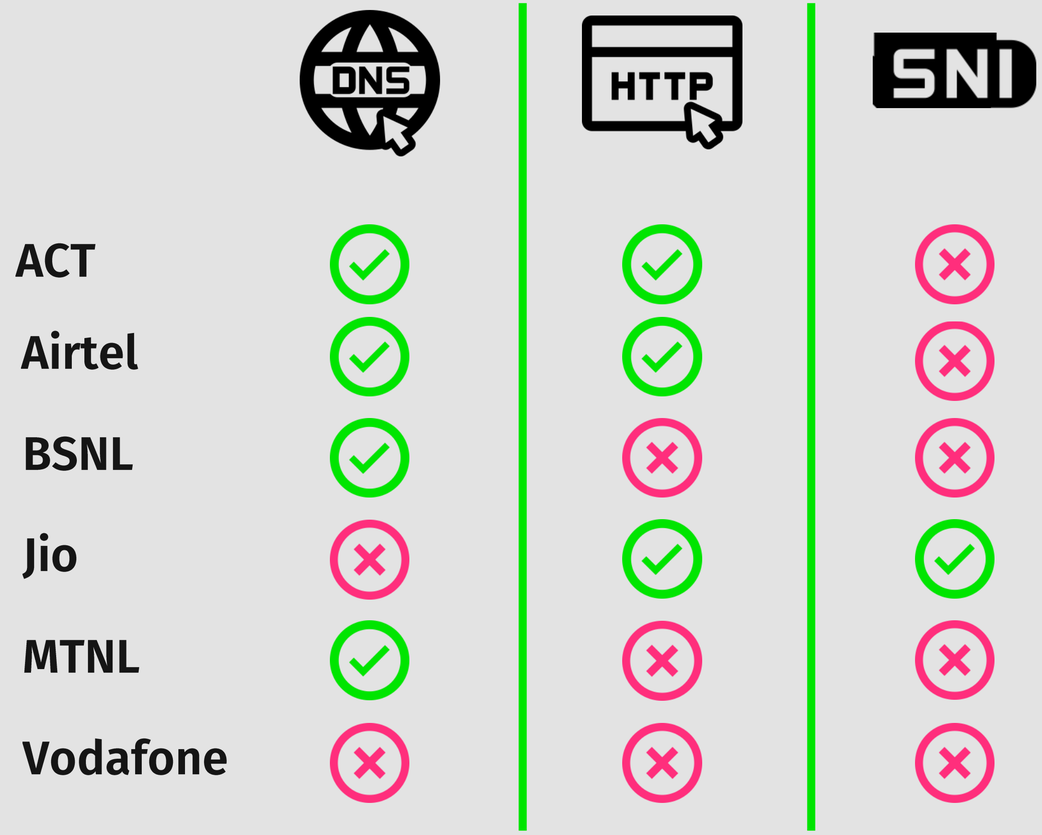
SNI

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



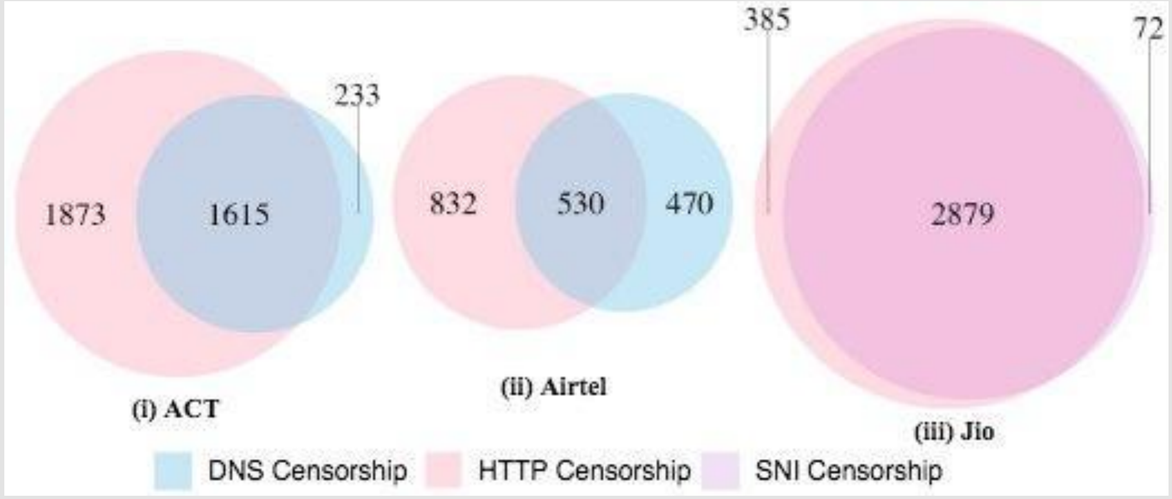
Different censorship mechanisms, individually or in combination to censor websites.

Results: Censorship Techniques

	DNS	HTTP	SNI
ACT	✓	✓	✗
Airtel	✓	✓	✗
BSNL	✓	✗	✗
Jio	✗	✓	✓
MTNL	✓	✗	✗
Vodafone	✗	✗	✗

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

- 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

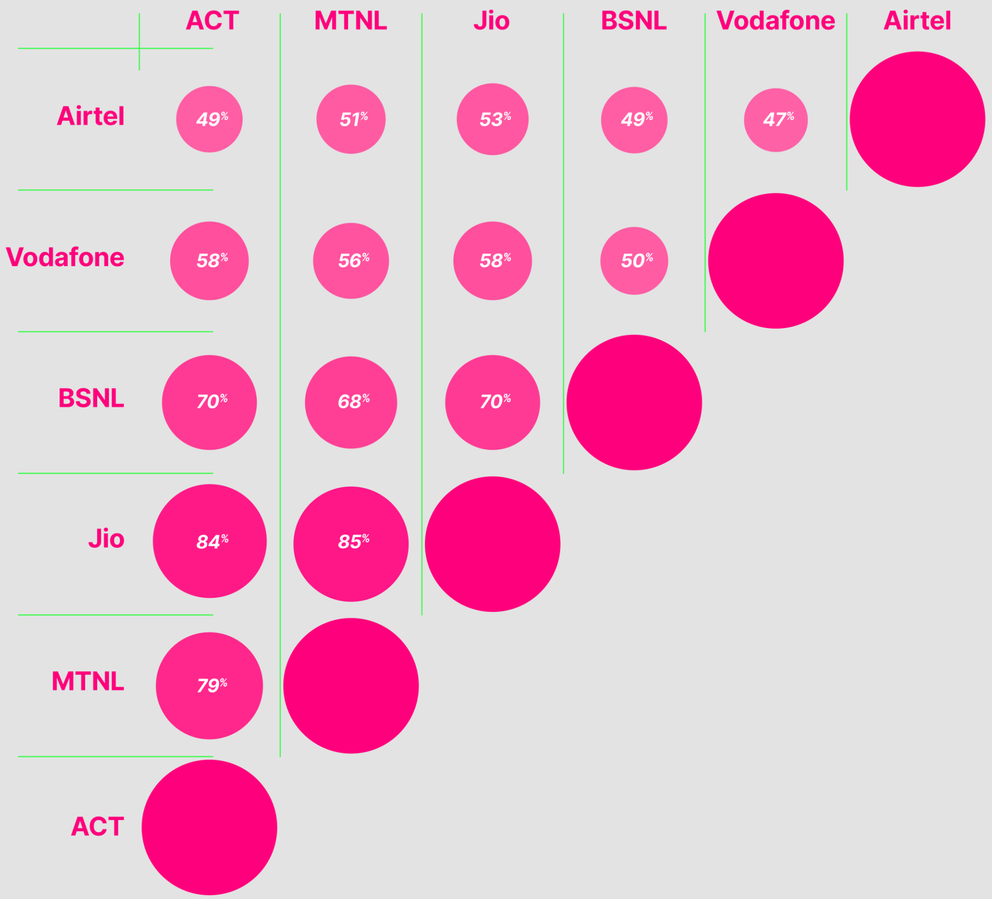
27.64%

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

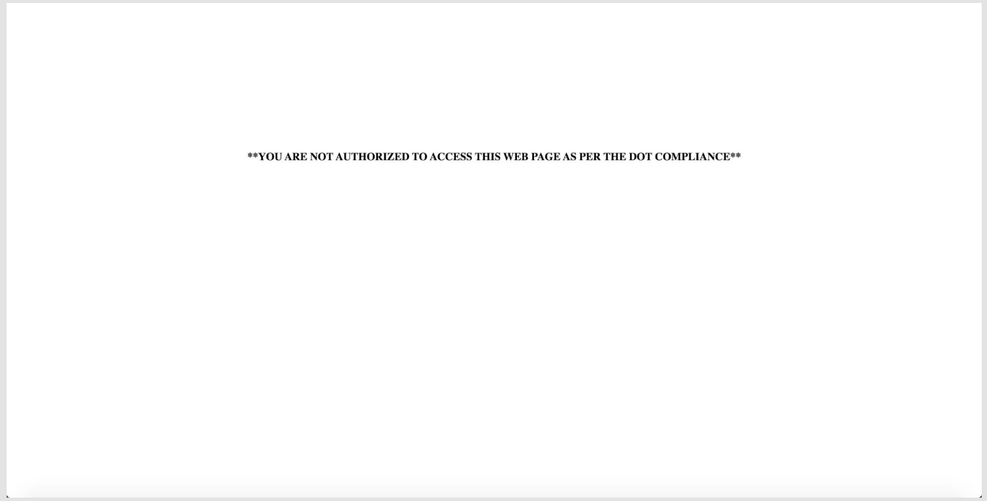
1. Need to re-evaluate legal and technical mechanisms of web censorship in India

2. Have a net neutrality monitoring mechanism in place

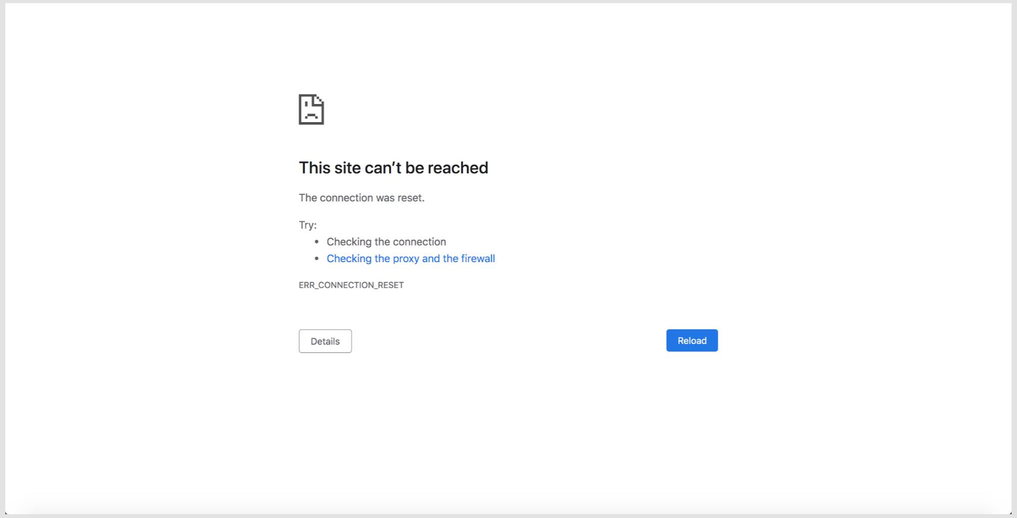
Conclusion

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

Future work

1. Efficient censorship circumvention

2. Get readings from all across the country (we’re working on a mobile app now)

QUESTIONS, FEEDBACK?

How India Censors the Web

Kushagra Singh, **Gurshabad Grover**, and Varun Bansal
(thanks to Akash Sheshadri for the slide designs!)

Email: gurshabad@cis-india.org