Manzar: Issues with Cyber Censorship in India

Student Name: Sahil Shekhawat
Roll Number: 2013083

BTP report submitted in partial fulfillment of the requirements
for the Degree of B.Tech. in Computer Science & Engineering

on 6th July 2017

BTP Track: Research

BTP Advisor
Dr. Sambuddho Chakravarty

Indraprastha Institute of Information Technology
New Delhi

Student's Declaration

I hereby declare that the work presented in the report entitled **"Manzar: Issues with Cyber Censorship in India"** submitted by me for the partial fulfillment of the requirements for the degree of *Bachelor of Technology* in *Computer Science & Engineering* at Indraprastha Institute of Information Technology, Delhi, is an authentic record of my work carried out under guidance of **Dr. Sambuddho Chakravarty**. Due acknowledgements have been given in the report to all material used. This work has not been submitted anywhere else for the reward of any other degree.

..... **Place & Date: .....**
Sahil Shekhawat

Certificate

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

..... **Place & Date: .....**
Dr. Sambuddho Chakravarty

Abstract

Last semester we found that the Indian government delegates censorship policies at ISP level which is inconsistent and disorganized. We filed multiple RTIs to Department of Telecommunication seeking censorship policies of Indian government but did not get a clear response. Then we tested various ISP and found evidences of same kind of censorship happening.

This semester we tested how censorship is happenning across various ISPs and found the mechanism to be same across ISPs, even the blocking message was same across multiple ISPs which leads to two possibilties, i.e. either every ISP is using the same tool or censorship is happening only at the ISPs who are at the top of the hierarchy e.g. Airtel. Present state of art tools like OONI [1] does not tries to either maintain a list of blocked websites or try to circumvent it. Furthermore, it does not have the functionality to locate middle-boxes which are doing it. We fixed all of the above shortcomings of OONI.

We created our own set of tests for censorship to determine type of censorship and locate middle-boxes. We designed a tool to circumvent the censorship happening in India. We are also working on a platform i.e. a web app and a chrome extension to collect result of our tests in a central repository to further aid future studies in this domain.

Keywords: Network-censorship, Surveillance, India.

Acknowledgments

I would like to most sincerely thank Dr. Sambuddho Chakravarty for being my advisor. His expertise, consistent guidance and ample time made this possible. I would also like to thank Devashish Gosain. Both of them steered the project in right direction with well structured guidance and input. They allowed me to own the project but also kept me on the correct path. I look forward to keep working with them.

Contents

Chapter 1

Introduction

Last semester we found that the Indian government delegates its policies to ISPs which in turn use another enterprise software to implement it. We wanted to answer following questions:

This semester we tried to map the exact working of censorship in India i.e.

- *How ISPs actually block our request? Do they hijack the TCP connection?*

- *Are all ISPs using same tool? If not, then how come we see same block message across multiple ISPs?*

- *Can we build a tool to test and circumvent it?*

- *How can we maintain an updated list of blocked websites?*

To answer the above questions we started by manually inspecting ISPs which were known to do censorship e.g. Airtel. we found that it was acting like a transparent proxy which is doing url filtering by hijacking the tcp connection between user and a blocked website. This type of hijacking is prone to errors and we were able to access blocked websites by sending the same request multiple times. List of blocked website was not consistent between ISPs and one ISP i.e Reliance JIO which was not doing any kind of censorship, implemented censorship during the course of experiments.

We now start with the Background and Related work

Chapter 2

Background and Related Work

Similar research has been done for openly censorious nations like China [2,3], Iran and Pakistan but not for India, which has an Internet population the size of Europe. India is therefore, significant to study censorship.

Also, the interaction of Internet with government policies is a well researched topic but most of them are focused on China whereas we focused on India.

OONI, the Open Observatory of Network Interference, is a similar attempt from Tor Project to collect high quality data using open methodologies, and Free and open source software to share observations and open data about various types, methods and amount of network tempering in the world including India.

Chapter 3

Our Approach

3.1 Finding government policies

We filed two RTIs to Department of Telecommunication ie. **DOTEL/R/2017/50126** and **DOTEL/R/2017/50133** asking following questions:

1. I would like to know the list of blocked websites/services in India. Please provide the list of websites which are blocked by Department of Telecommunication in India.
2. Please specify if the websites are blocked in all over India or in few specific states.
3. Also, please specify if websites are blocked by few ISPs or by all ISPs in India.

And the response we got was:

1. *Internet service licenses are to follow the provisions of Information Technology Act 2000 as amended from time to time. Under Information Technology Act 2000, Information Technology (Procedure and Safeguards for blocking for Access of Information by public) Rules 2009 were notified on 27/10/2009. Aforesaid notified rules describes the Designated officer for the purpose of issuing direction for blocking. Vide Gazette Notification dated 20/01/2010, Group Coordinator, Cyber Law Division, Department of Electronics and Information Technology has been authorized and designated as Designated Officer. The said IT Act, Rules and the Notification are available at the website of Department of Electronics and Information Technology, that is, www.deity.gov.in thereof in few cases, based on the court direction, action is initiated by DeitY.*
2. *Further, Clause 16 of Information Technology Procedure and Safeguards for Blocking for Access of Information by Public Rules 2009 says that Strict confidentiality shall be maintained regarding all the requests and complaints received and actions taken thereof.*
3. *Accordingly, as per above mentioned rules, the directions for blocking of websites for access by public is Issued by Group Coordinator, Cyber Law Division, Department of Electronics*

and Information Technology (DeitY), Electronics Niketan, CGO Complex, New Delhi. As per the direction of Group Coordinator, Cyber Law Division, under Information Technology Act 2000, Instruction for blocking/unblocking of websites/URLs are issues to all Internet Service Licensees of DoT. In few cases, based on the court direction, action is/DoT.

which is a very diplomatic response without providing any kind of useful information. This goes on to prove how inconsistent censorship in India really is because there is no central organization which controls it.

Since, we were not able to get a list of blocked websites from government, we compiled our own list from various sources including websites which were proven to blocked at some point of time.

3.2 Manually inspecting ISPs

3.2.1 What type of packets are being blocked?

We ,using our list of blocked websites, manually checked how the request was being blocked using Wireshark and found evidences which goes on to prove our hypothesis. We used following approach on Airtel (largest ISP in India) on a particular url *www.wapmaza.mobi* known to be blocked.

1. **Traceroute**: To see if all packets to a blocked destination (including ICMP) were being blocked and found that they were not. We were able to reach desired destination using ICMP packets.

2. **TCP Traceroute**: To see if other type of packets were being blocked i.e. TCP packets and found that they were not. TCP Traceroute was able to reach desired destination similar to Tracerout e.

3. **HTTP Requests**: Finally, we saw if HTTP Requests were being blocked and they were. We were receiving a message from Airtel i.e. *Your requested URL has been blocked as per the directions received from Department of Telecommunications, Government of India. Please contact administrator for more information..* Now, we confirmed that, that specific url was being blocked.

3.2.2 How ISPs implements blocking?

Using Wireshark we found that we are getting a first HTTP response packet directly after our request which our browser treats as the response and renders it. We also received a different HTTP response packet with same sequence number as first HTTP response packet from the same source. We found, using Tor, that the HTML in the second packet was same as the actual website but second packet was counted as a retransmission and therefore, dropped. Now, we

knew that first packet was from ISP and second was from the actual server and needed to prove it further.

To confirm that the first packet was not from blocked website's server we compared epoch timestamps on the HTTP request and response and compared the RTT (Round Trip Time) of HTTP packet with RTT of Traceroute. We found:

1. RTT of first HTTP packet was much smaller that RTT of Traceroute.
2. RTT of second HTTP packet was comparable to Traceroute's RTT.

3.2.3 Finding ISP responsible for blocking

To see if all ISPs were doing censorship, we manually tried on a very small ISP *Broadsol* and found Airtel's blocking message on *www.wapmaza.mobi*. Now, we needed to prove is Airtel itself doing censorship or is it serving its blocking page to this small customer ISP.

To solve it, we calculated the number of hops between customer and server using tcptraceroute, then we build a tool to consistently lower TTL starting from the number of hops to 1 . We found that even if TTL is lower than the number of hops we received the HTTP response containing the blocked message till a certain TTL. After that we stopped receiving any response. This leads to the following conclusions:

1. ISP is doing censorship by deploying middle-boxes within its network which capture the traffic based on HTTP request. i.e. value of HTTP Header 'Host'.
2. Middle-box checks every HTTP request, i.e. no IP filtering and responds with the blocked message. It does not check anything else.

But, above was not true for all requests we sent. Therefore, we checked whether it is possible for the second HTTP packet (actual response from website) to reach user before the first packet.

3.2.4 Is blocked website always blocked?

We sent 100 curl requests with same HTTP headers immitating a web browser and noted number of times we received Airtel's blocked message and number of times we received an error.

Number of times website was unblocked = 100 - (Number of times we received blocking message + Number of times error occurred.)

www.wapmaza.mobi was blocked 88 times with 0 erros. Therefore, it was unblocked 12 times.

12% successful requests shows that censorship in India is not only inconsistent across ISPs but within an ISP too in spite of the fact that they deploy similar machanism.

3.3 Running OONI

After confirming that censorship was indeed happening in India, we used OONI. Tests we used for confirming blocked websites were:

1. **Web connectivity:** This test examines whether websites are reachable and if they are not, it attempts to determine whether access to them is blocked through DNS tampering, TCP connection RST/IP blocking or by a transparent HTTP proxy.
2. **HTTP Host:** This test attempts to:
   (a) Examine whether the domain names of websites are blocked
   (b) Detect the presence of middle boxes (software which could be used for censorship and/or traffic manipulation) in tested networks
   (c) Assess which censorship circumvention techniques are capable of bypassing the censorship implemented by the middle box
3. **HTTP Requests:** This test tries to detect online censorship based on a comparison of HTTP requests over Tor and over the network of the user.

To find out the type of censorship that was happening, we used:

1. **HTTP Header Field Manipulation Test:** This test tries to detect the presence of network components (middle box) which could be responsible for censorship and/or traffic manipulation.
2. **HTTP Invalid Request Line:** This test tries to detect the presence of network components (middle box) which could be responsible for censorship and/or traffic manipulation.

3.4 Running our own tests

To make a comparison with OONI's results, We wrote our own tests which:

1. **Test website reachability multiple times:** We sent 100 requests for each website in our link from multiple ISPs to test number of times, we were successfully able to get the correct response.
2. **Test whether all ISPs were using same mechanism:** We earlier found that ISPs were not stopping requests to blocked website's server but inturn sending their own response immitating the actual server. We wanted to find whether this holds true for all ISPs.
3. **TTL test** By continuously reducing TTL by 1 and looking for HTTP response gave us a clear example of middle-box censorship in India.

Chapter 4

Experimental Results

Below are some initial results. More experiments are still running and will have a lot of more results comparing different ISPs and different tests.

ISP	**Website Categories**							
	Escort (150)	**Music** (100)	**Porn** (50)	**Torrents** (30)	**Social** (20)	**Political** (20)	**Tools** (20)	**Misc.** (150)
Airtel	50, 80, 20	82, 6, 12	1, 49, 0	13, 16, 1	8, 10, 2	2, 15, 3	1, 14, 5	80, 41, 29
Vodafone	24, 87, 39	95, 1, 4	2, 45, 3	16, 11, 3	8, 8, 4	0, 13, 7	4, 11, 5	70, 35, 45
Sify	12, 98, 40	1, 75, 24	1, 48, 1	6, 22, 2	0, 16, 4	0, 15, 5	1, 16, 3	11, 75, 64
NKN	11, 105, 34	57, 33, 10	1, 48, 1	10, 16, 4	4, 12, 4	2, 14, 4	1, 14, 5	65, 56, 29
BSNL	41, 69, 40	68, 12, 20	0, 45, 5	12, 14, 4	7, 10, 3	4, 12, 4	3, 14, 3	88, 27, 35
MTNL	27, 98, 25	81, 2, 17	45, 3, 2	15, 12, 3	9, 8, 3	14, 1, 5	2, 12, 6	73, 23, 54
Siti	23, 99, 28	28, 56, 16	44, 4, 2	14, 13, 3	9, 8, 3	1, 14, 5	1, 12, 7	86, 29, 35
Reliance Jio	0, 123, 27	0, 77, 23	0, 38, 12	2, 26, 2	0, 18, 2	0, 16, 4	0, 15, 5	0, 78, 72

Table 4.1: Censorship trends in India: Some initial results

Table 4.2: Some of the results showing partial blocking of websites

websites	blocked requests	errored requests
www.wapindia.net	90	5
3gparena.in	46	0
uptobox.com	95	0
radiorhythmz.fm	81	0
apunkabollywood.com	93	0
bollywoodmp4.com	93	0
nowvideo.sx	93	7
mypetjawa.mu.nu	94	0
extratorrent.cc	66	0
tamilmaalai.com	87	13
www.radiorhythmz.fm	84	0
mhnwap.wapka.mobi	87	0
krazywap.com	90	0
riyamodel.in	90	0
nowvideo.ch	96	0
mp3feelings.com	93	7
www.mumbaidolls.com	89	1
www.asika.in	86	14
netmasty.com	89	0
uploadboy.com	89	0
mimti1.moviesmobile.net	90	10
arabloads.net	96	0
h33t.to	91	0
shneha.in	93	7
www.wapindia.net	90	5
3gparena.in	46	0
uptobox.com	95	0
radiorhythmz.fm	81	0
apunkabollywood.com	93	0
bollywoodmp4.com	93	0
nowvideo.sx	93	7
mypetjawa.mu.nu	94	0
extratorrent.cc	66	0
tamilmaalai.com	87	13

Table 4.3: Reliance JIO

websites	total Hops	blocking at/after hops
apunkabollywood.com	26	10
bitvid.sx	21	10
radiorhythmz.fm	21	10
nowvideo.sx	24	10
bollywoodmp4.com	23	10
ojaloberoi.in	25	10
wapindia.net	26	10

Table 4.4: NKN

websites	total Hops	blocking at/after hops
apunkabollywood.com	14	8
bitvid.sx	6	8
radiorhythmz.fm	11	8
nowvideo.sx	16	8
bollywoodmp4.com	14	8
ojaloberoi.in	16	8
wapindia.net	22	8

Table 4.5: Sify

websites	total Hops	blocking at/after hops
apunkabollywood.com	27	10
bitvid.sx	23	10
radiorhythmz.fm	23	10
nowvideo.sx	26	10
bollywoodmp4.com	25	10
ojaloberoi.in	27	10
wapindia.net	28	10

Chapter 5

Limitations and Future work

The experiments are not finished and we are running the same tests on more ISPs to get a more clear picture of censorship in India. Some of the limitations are as follows:

1. Access to limited number of ISPs in India: There are 247 registered ISPs in India and reaching a good chunk of them covering all states and all types of ISPs is very difficult.
2. High number of False positives in OONI: We did created our own tests but they were not as extensive as OONI's but it is documented that OONI produces a high number of false positives.
3. No open documentations for enterprise censorship products like **Netsweeper**: which is used in many countried like Qatar and might be used in India as well.
4. Lack of an updated list of blocked websites in India.

Chapter 6

Concluding Remarks

We are still working on the platform for a centralized repository of blocked websites . We are building a chrome extension to automatically determine blocked websites a user is facing and update the central repository about the blocked website, country of user and his ISP.

Inspite of some big and important limitations we are able to do some key findings about Censorship in India. We are also working to improve OONI and provide all of our results and data as open data.

Bibliography

[1] OONI. `https://ooni.torproject.org/`.

[2] Park, J. C., and Crandall, J. R. Empirical study of a national-scale distributed intrusion detection system: Backbone-level filtering of html responses in china. In *Distributed Computing Systems (ICDCS), 2010 IEEE 30th International Conference on* (2010), pp. 315–326.

[3] Winter, P., and Lindskog, S. How the Great Firewall of China is blocking Tor. In *Proceedings of the USENIX Workshop on Free and Open Communications on the Internet (FOCI 2012)* (August 2012).