Russian Collateral Censorship

Jesse Brizzi
jbrizzi@cs.stonybrook.edu

Konstantin Dmitriev
kdmitriev@cs.stonybrook.edu

Yingtao Tian
yittian@cs.stonybrook.edu

***Abstract*—In this project we are proposing a survey of neighboring sovereign nations to modern day Russia to investigate any possible collateral censorship. Given the socioeconomic state of some of these countries there may be limited infrastructure for Internet access, which requires their traffic to be routed through Russian territory. Russia actively censors various websites and Internet sources based on various reasons.**

Keywords*—*Collateral Censorship, Russia, Block, DPI, DNS, IP.

I. Motivation

In the summer of 2008, when Russia's mass media and telecom watchdog Roskomnadzor [1] was re-established, the Russian Internet, or RuNet, changed significantly. This Federal Service is regulated and put into motion by two laws, - "On Protecting Children from Information Harmful to Their Health and Development"[2] and "On Information, Information Technology and Information Protection"[3]. Both of them give judges a free hand in decision-making. As a result, a number of websites have been blocked quite chaotically, starting with opposition websites and articles, to Bitcoin communities and GitHub.

Sometimes such censorship systems can case collateral censorship, or damage. They block access to sites from users beyond those intended to protect[1]. This projects goal is to examine the effect of such possible collateral censorship to the requests that are originating from outside of Russia, with the possible extension to other countries that maintain the censoring services. The result of this project can be used to create a detailed analysis of collateral damage caused by different types of censoring techniques, and to potentially discover the paths at fault.

II. Related Work

Not a lot of research has been done in the area of the collateral censorship between networks in different countries. Partly because the impact of Internet censorship on global Internet service is usually unintended, and the probability of getting any results is fairly small. However, Chinas injection of forged DNS responses has been reported to cause large scale collateral damage by blocking outside traffic that traverses Chinese links [1]. The analysis shows that in the most extreme case, 70% of the open resolvers from Korea suffer collateral damage for queries to .de domains. Upstream filtering can also be behind traffic blockage outside of a censoring area due to ISP routing arrangements (for example, the Indian Internet filtering some users in Oman who are not able to access certain webpages [4]).

[1]Federal Service for Supervision of Communications, Information Technology and Mass Media (Russian:)

III. Censorship tools

The Russian Internet Service Providers (ISPs) and government use a number of different censorship techniques to block access to "unwanted" websites. Fig. 1 shows a chart of the most popular ones by the number of providers that maintain a particular method.

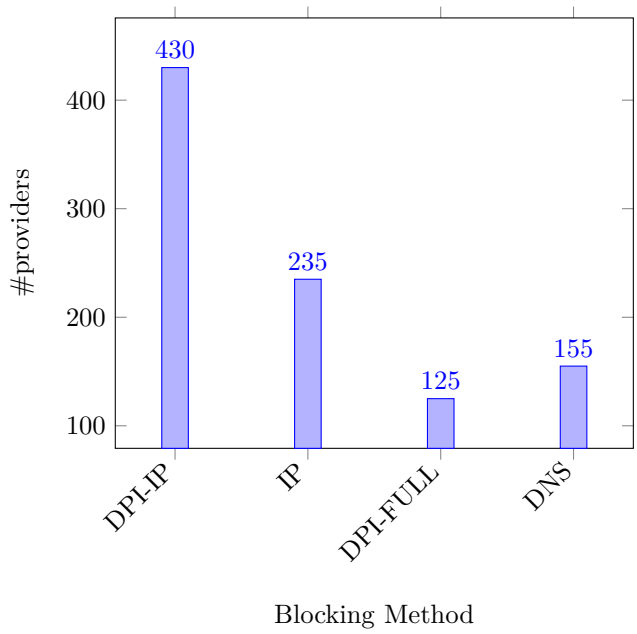

Figure 1: Number of providers that maintain a particular blocking method (DPI-IP - blocking using Deep Packet Inspection (DPI), that checks URL only at a specific IPs and port:80; DPI-FULL - blocking using DPI at every IP and ports; IP - blocking by IP address; DNS - DNS injection)

IV. Resources

As a main reference for blacklisted URLs, domain names, and IPs in Russia, we plan to use the service http://antizapret.info which maintains and persistently updates its catalog. In addition, it also provides an API for external access.

Listing 1: An example of the possible output from requesting http://api.antizapret.info/all.php?type=json

```
{
"id":"16971",
"rsoc_id":"91191193",
"includeTime":"2015-02-13 11:55:47",
"rsocDate":"2015-02-04",
"org":"",
"org_act":"2/1/11-32425",
"url":http://chetkiibro.com/,
"domain":"http://chetkiibro.com/,
"ip":"104.28.16.89,104.28.17.89",
"country":"",
"proof":"http://antizapret.info/site.php?id=16971
    "
```

As a reference of the potential victims of the collateral censorship, we are going to use telegraphy maps (Fig. 2), and the map of the supported countries for one of the largest Russian backbone service provider (Fig. 3), concentrating on the bordering countries.

Also we are using VPNs provided by IP Vanish [2]

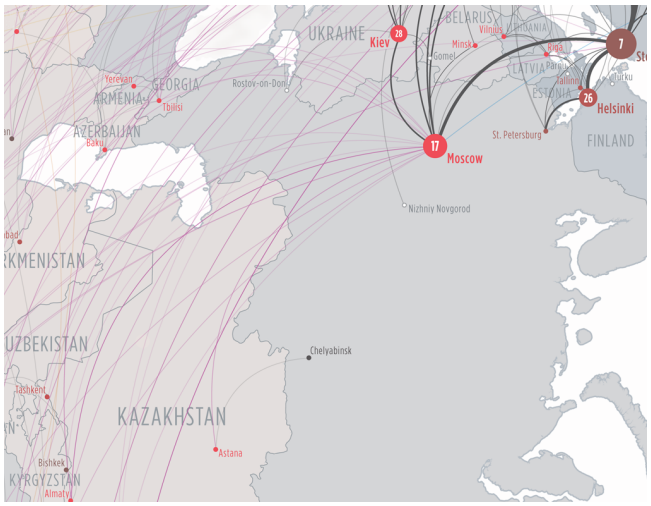

Figure 2: Telegeography Map of the Region

Figure 3: RETN's backbone map

[2]https://www.ipvanish.com/

V. Research Plan

A. Ultimate Plan

Our ultimate aim is to write an experiment in the form of Python scripts that will probe websites censored in Russia from various points in neighboring countries. This experiment will not only be able to compare the received web-pages with the blocked ones, but also return the type of the applied censorship technique in case of detected similarity between the received and the blocked web-pages. To collect the source code of blocked web-pages we plan to use our groups own private computer in Russia or a VPN service

In order to create this experiment, we plan to use the following modules:

1) **urllib** - for network resource access;
2) **socket** - to get an access to the BSD socket interface;
3) **ssl** - to get an access to Transport Layer Security encryption and peer authentication facilities for network socket;
4) **dnspython** - to get an access to high and low levels of DNS.

and, of course, the OpenVPN [3] client to connect to VPN servers. Also we utilize the scripts from Tunnelblick [4] for connection setup.

The coarse description of the experiment is the following:

1) Collect a list of URLs and IPs of the censored web-pages from http://antizapret.info;
2) For each item in the list:
   a) Request the web-page and compare its source code with the code of actually blocked web-page;
   b) Test DNS access;
   c) Test HTTP access.

B. Contingency Plan

Given the possibility where no censorship leakage is found, we will restructure the experiment to try and prove our results are accurate, i.e. that there is little to no collateral censorship resulting from the Russian government. This will be done by repeating the experiment multiple times at different times of the day, along with expanding our list of candidates for possible collateral censorship to other countries in close proximity geographically and in terms of Internet topology.

VI. Methodology

In order to measure possible collateral damage caused by Russian censorship, we conduct an experiment that can be coarsely divided into 3 parts:

1) Set a connection to a VPN server from a list of examined countries.

[3]http://openvpn.net/
[4]https://code.google.com/p/tunnelblick/

2) Access and collect HTML code of web-pages that are blocked in Russia.
3) Compare the percent of similarity between web-pages access from Russia and outside of it.

The overall process is illustrated in Figure 4.

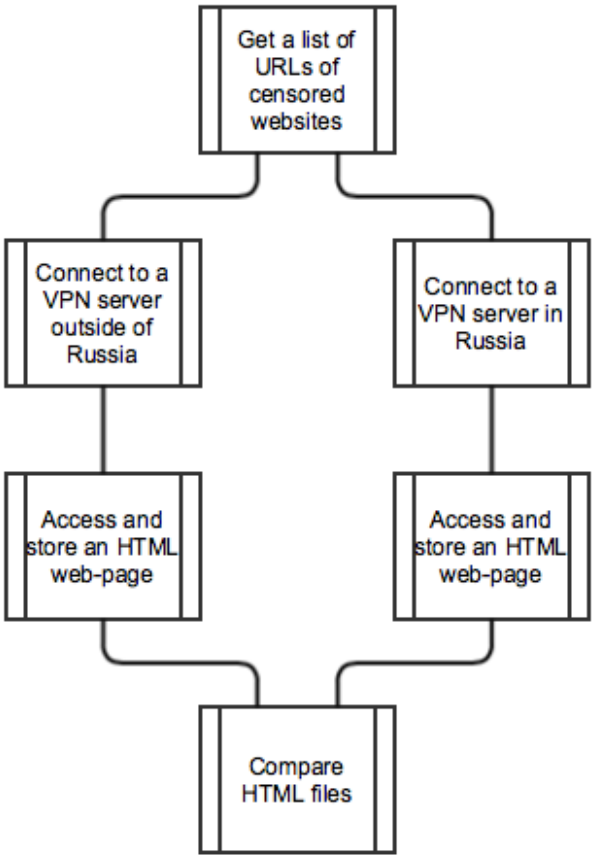

Figure 4: Scheme of the experiment

A. VPN

We are using servers provided by IP Vanish. According to it's website [5], its servers " span 25,000+ IPs on 165+ servers in 60+ countries". This is a huge advantage as countries near Russia, like Finland and Estonia, are among countries supported by IP Vanish, thus we can investigate the influence of Russian censorship on its geographical neighbors. Also, we are using a computer directly located in Russia to get more accurate result for Russia censorship.

For our VPN client we are using OpenVPN. Basically, we conntect to a OpenVPN server hosted by IP Vanish using configuration files provided by IP Vanish and associated credits. Also we are using scripts from Tunnelblick to handle setting-up / tearing-down the connection on Mac OS X systems. We are running OpenVPN in daemon mode, communicating with it via a telnet server open locally for management purpose.

[5] https://www.ipvanish.com/why-vpn.php

B. Data Collection

After setting up a connection to a VPN server in a particular country, we access web-pages from a list of censored websites in Russia. To access a this list we use an API from **AntiZapret**.

Basically, going through every URL in that list we access and store the respective HTML web-page from each country.

For the data collection purpose we implemented a special class - *CollectData*. Its instance connects to a specified VPN server and performs either collection of a single wep-page, specified by a URL from multiple countries, or collects a list of web-pages, using single VPN connection (Country major vs Website major order). This gives us the option of targeting a specific website in a small time windows to minimise differences in time sensitive content.

After the web-page access attempt, there could be four possible outcomes:

1) An original web-page.
2) A web-page with a removed part (partial censorship).
3) A censorship's system brochure web-page.
4) A failure to load a web-page.

In order to deal with the final case, we stop waiting for a response after 8 seconds (Timeout parameter).

After the data collection procedure, we perform a string by string comparison between web-pages accessed from a VPN server located in Russia, and web-pages accessed from a VPN servers located in a selected list of countries. The result of such comparison is a percentage of similarity between pages.

VII. Web User Interface

To display the results of our tests we will create a Web UI, that will consists of a table with the outcomes of the particular experiment, including URL addresses, URL categories, statuses and explanations. That Web UI will also have an option that will give the user a visual summary for the all conducted experiments.

VIII. Timeline

Estimation of progress by week.

2/16 - 2/22 ~~Setup and design~~
2/23 - 3/01 ~~Design Interface Website/Experiment~~
3/02 - 3/08 ~~Code Interface Website/Experiment~~
3/09 - 3/15 ~~Code Interface Website/Experiment~~
3/16 - 3/22 ~~Spring Break/Coding/Debugging~~
3/23 - 3/29 ~~Run Experiment~~
3/30 - 4/05 ~~Run Experiment/Analysis Data~~
4/06 - 4/12 ~~Midterm project Report due~~
4/13 - 4/19 Expand Experiment if needed
4/20 - 4/26 Create Website for Results
4/27 - 5/03 Buffer Week
5/04 - 5/10 Final Project Report Due

IX. Current Issues

A major obstacle that we are working to overcome is the inconsistency of the censorship from our different viewpoints from *within* Russian borders. Using our various VPN subscriptions and a physical computer that we have access to in Russia we are observing different levels of access to certain websites. Our PC in Russia is censored in more cases then it is not when testing against the list of target websites. Where our VPN service through IP Vanish seems to get through to most of these websites. This may be due to the VPN company purposely fetching these locked websites as this would be a feature that most of its users would want or their servers are falsely labeled as being located in Moscow. We just happen to be in the fringe case of actually wanting to be censored.

We plan on either only collecting Russian samples from our dedicated PC or trying a difference VPN service. An issue may be similar de-censorship in the neighboring countries too, but we do not have dedicated machines in these locations to check. This may lead to a lot of invalid results in the end.

X. Future Work

Currently we plan on focusing on expanding our HTML comparison methods to try and find what type of things are being changed if only parts of webpages are being censored. To account for different dynamic content that may change from instance to instance (ads, time dependant information).

We have references to methods of figure printing the type of censorship being used from the Russian ISPs, we would like to integrate this into our project as well.

Finally we still want to create the website to display our results pending we find proof that is worthy of being published outside of the class.

References

[1] Anonymous, *The Collateral Damage of Internet Censorship by DNS Injection.* SIGCOMM Comput. Commun. Rev., July 2012.

[2] *"Law on Protecting Children from Negative and Harmful Information."* President of Russia. N.p., n.d. Web. 14 Feb. 2015.

[3] *Russian Federation: Federal Law No. 149-FZ* of July 24, 2006, on Information, Information Technology and Information Protection (as Amended up to Federal Law No. 398-FZ of December 28, 2013). N.p., n.d. Web. 14 Feb. 2015.

[4] C. Lab., *Routing Gone Wild: Documenting upstream filtering in Oman via India.* Technical report, Citizen Lab, 2012.