Decentralized Control
A Case Study of Russia

Reethika Ramesh, R. Sundara Raman, M. Bernhard, V. Ongkowijaya, L. Evdokimov, A. Edmundson, S. Sprecher, M. Ikram, and R. Ensafi

24 February 2020
Centralized Censorship

- Conventionally, censorship = centralized
  - China developing the GFW over the past 17 years
  - High investment in money and time
Decentralized Censorship Infrastructure

- Multiple ISPs with different motivations
- From a govt perspective:
  - Synchronizing policies
  - Large scale
  - Real time filtering
- Russia has been ramping up: despite 1000s of ASes
Russia’s Model: Decentralized Censorship Apparatus

- Russia is building their national censorship apparatus
- Facilitated by the commoditization of filtering technologies
- From a research standpoint:
  - Is decentralized censorship feasible to implement?
  - How effective is it?
  - Can other nations adopt it easily?

→ Need to conduct meaningful measurements
Censorship Measurement Checklist

1. Identifying domains to test
2. Diverse vantage points
3. Sound control measurements
Identifying Domains to Test

- Worked extensively with activists
- **Obtained 5 leaked digitally signed samples** of authoritative blocklist
- Pointed to repository that tracked the leaked blocklist over time
  ➔ Found 99% similarity between signed samples and repository entries

Signatures use GOST CN=Роскомнадзор or CN=Единая информационная система Роскомнадзора (RSOC01001), translates to “Roskomnadzor,” and “Unified Information System of Roskomnadzor.”
Characterizing the Blocklist

We characterized:

➔ 7 years worth of historical data with commits of daily granularity
➔ Rapid growth

132,798 Domains
324,695 IPs
39 Subnets
Characterizing the Blocklist

- 63% websites had content in Russian, 28% in English
- State of the art categorization services don’t work well for languages other than English
- Developed our own topic modeling algorithm
Topic Modeling

1. Text Extraction - Used **Beautiful Soup** to extract text from HTML
2. Language Identification - Python’s **langdetect** library

Ran the rest for Russian and English separately
3. Stemming - Reduce words to stems using Snowball
4. TF-IDF - Term frequency-inverse document frequency
5. LDA analysis - Python’s **gensim** and **nltk**

→ Arrived at 20 topic word vectors each for English and Russian, then labelled manually
Characterizing the Blocklist

➔ **Popular categories** were gambling and pornography, also:
  ○ Russian news websites with political content
  ○ Circumvention websites
Censorship Measurement Checklist

1. Identifying domains to test
2. Diverse vantage points
3. Sound control measurements
- Rented 6 VPSes
- Recruited 14 participants to run residential probes
  - Ethically with informed, explicit consent
- To obtain a holistic view, we obtained vantage points to run remote measurements
Censorship Measurement Checklist

1. Identifying domains to test
2. Diverse vantage points
3. Sound control measurements
Sound Control Measurements

- Prune away the domains and IPs that are non-responsive
- 13 geographically distributed control vantage points
- Resolved all domains and made HTTP GET requests
- Made TCP connections to port 80 to all IPs in list and subnets

98,098 Domains
121,025 IP Addresses
31 Subnets
<table>
<thead>
<tr>
<th>#</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TCP/IP Blocking</td>
</tr>
<tr>
<td>2</td>
<td>DNS Manipulation</td>
</tr>
<tr>
<td>3</td>
<td>Keyword Based</td>
</tr>
</tbody>
</table>
Conducting Measurements

Direct Measurement
From datacenter VPSes and residential probes
- In-depth measurement
- Limited scale

Remote Measurement
From the remote measurement vantage points
- Large scale measurements
- Helps corroborate results for domains on the list
Conducting Direct Measurements

Local DNS Resolver

DNS Manipulation

GET a.b.c.d

VPS/Probe

domain.com

a.b.c.d

a.b.c.d
Conducting Direct Measurements

Keyword Based Manipulation

VPS/Probe

GET domain.com

domain.com
Conducting Direct Measurements

IPs in List and Subnet

VPS/Probe

TCP SYN to Port 80

a.b.c.d
Conducting Remote Measurements

- Ran remote measurements using **Quack** and **Satellite** to corroborate results
- Over 1000 vantage points in total
This is the first comprehensive, in-depth study that:

➔ uses an **authoritative blocklist** to investigate feasibility of decentralized information control and,

➔ combines views from **data centers, residential, and remote vantage points** to obtain a holistic view of censorship in a country.
Results

→ Domains (Direct and Remote)
→ IPs and Subnets (Direct)
Measurement Results for Domains

- Residential probes observe **high** level of blocking
- Significant difference in both **types and amount** of blocking between data center and residential vantage points
- Residential ISPs are more likely to inject **informative blockpages**
Measurement Results for Domains

- Only few data center VPSes observe blocking
- Data center networks less likely to inject blockpages, instead use resets and timeouts
- Residential ISPs:
  - Inject notices citing the law in blockpages
  - Sometimes even include advertisements!
Доступ к информационному ресурсу ограничен на основании Федерального закона от 27 июля 2006 №149-ФЗ «Об информации, информационных технологиях и о защите информации».

Ой!
Доступ к информационному ресурсу ограничен на основании Федерального закона от 27 июля 2006 г. №149-ФЗ "Об информации, информационных технологиях и о защите информации."
Узнать причину."
Remote Measurements Results

Fraction of domains blocked at the individual vantage point as well as AS (aggregated) level

- The similarity between the lines shows that blocking is happening at the AS level.
- Our measurements using Satellite observed much more blocking compared to Quack measurements.
Remote Measurements Results

- Policies of blocking are carried out at the AS level
  - High similarity of blocking
- Confirms DNS manipulation in cases where
  - Most domains resolve to the same IP and that IP hosts a blockpage
Results for IPs and Subnets

- Overall for IPs, lesser blocking compared to domains
- Residential ISPs more likely to block domains than IPs
- Different ISPs may prioritize blocking different subnets
Censorship Measurement Checklist

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
</table>
| 1 | Identifying domains to test  
Working with activists enabled us to obtain an authoritative test list |
| 2 | Diverse vantage points  
Obtained data center, residential, and remote vantage points to get a comprehensive picture of censorship in the country. |
| 3 | Sound control measurements  
Need strong controls to differentiate censorship from other failures |
Decentralized Control is Effective!

Our study finds:

- Implementing effective decentralized information control is feasible
- Commoditization of censorship & surveillance technology allows for simple solution
- Russia is succeeding at building a national censorship apparatus
Spreading Censorship Trends

**United Kingdom** - Government providing ISPs a list of websites to block and having governing censorship bodies that correspond to various types of censored material

**Indonesia** - Implementing content filtering at its network borders

**India** - has been ramping up censorship using Supreme Court orders imposed on ISPs

**United States** - the repeal of net neutrality is allowing ISPs to favor certain content over others
Spreading Censorship Trends

- Report in 2019 found Russian information controls being exported to 28 countries
- Enforce **accountability and transparency**
- Need **mechanism for auditing**
- Need **empirical, data-driven** studies to inspire change
Summary

- Highlight censorship measurement complexities
- Combine perspectives from diverse vantage points
- Prove that decentralized censorship is effective
- Illustrate impact of the use of commoditized technology for censorship
Decentralized Control
A Case Study of Russia

Reethika Ramesh, R. Sundara Raman, M. Bernhard, V. Ongkowijaya, L. Evdokimov, A. Edmundson, S. Sprecher, M. Ikram, and R. Ensafi

24 February 2020