Internet censorship

Tom Ristenpart
CS 6431
Current estimates of Internet censorship

OpenNet Initiative (ONI), Reporters Without Borders

Magenta-colored countries are countries with pervasive censorship and surveillance of the Internet

"Internet Censorship and Surveillance World Map" by Jeffrey Ogden (W163)
How would you censor web requests?

- IP filtering
- DNS filtering / redirection
- URL filtering
- Packet filtering (search keywords in TCP packets)
- Protocol filtering (detect Tor protocol)
Golden Shield Project

“If you open the window for fresh air, you have to expect some flies to blow in” – Deng Xiaoping in 1980s

Great Firewall of China:
• IP filtering
• DNS filtering / redirection
• URL filtering
• Packet filtering (search keywords in TCP packets)
• Protocol filtering
• Active probing of suspect destination IP addresses
Islamic Republic of Iran

• Every ISP must run “content-control software”
  – SmartFilter (up until 2009) made by USA company
  – Nokia Siemens deep-packet inspection (DPI) systems

• According to wikipedia: 50% of top 500 most popular websites blocked in Iran

• Occasional widespread filtering of Tor, TLS, other encrypted protocols
Censorship as two-step process

1. Sensitive content identification
   – DNS and IP blacklists
   – Keyword blacklists with DPI
   – Protocol identification (e.g., TLS)
   – Tool identification (e.g., Tor)

2. Censoring action
   – DNS poisoning
   – HTTP man-in-the-middle
   – TCP resets
   – Dropping packets
Types of packet inspection

Internet service providers need only look at IP headers.

Deep packet inspection (DPI) analyzes application headers and data.
DPI technology

• From Narus’ website (http://narus.com/index.php/product/narusinsight-intercept):
  – “Target by phone number, URI, email account, user name, keyword, protocol, application and more”, “Service- and network agnostic”, “IPV 6 ready”
  – Collects at wire speeds beyond 10 Gbps

• Narus allegedly used by NSA in San Francisco AT&T office
NarusInsight™ Selected To Save Pakistan's Telecommunications Networks
Millions Of Dollars Per Year

NarusInsight™ Selected to Save Pakistan's Telecommunications Networks Millions of Dollars Per Year

Narus System Chosen to Detect Rogue VoIP Traffic

MOUNTAIN VIEW, Calif.—September 21, 2007—Narus, Inc., the leader in carrier-class security for the world’s largest IP networks, today announced that the company has teamed up with Inbox Business Technologies Pvt. Ltd., a leading total IT solution provider in Pakistan, to keep Pakistan’s telecommunication networks clear of illegal, rogue and malicious IP traffic. NarusInsight was chosen by the Pakistan Telecommunication Authority (PTA) (the government administration responsible for regulating the establishment, operation and maintenance of telecommunication systems, and the provision of telecom services) to detect rogue VoIP traffic flowing through the telecommunications network in Pakistan.


Please come back later.
How do we know about censorship?

• Anecdotes from people within censored regions
• More formal surveys
• Network measurements:
  – Web sites aggregating info such as GreatFire
  – Herdict tool (browser plugin to manually report blockage)
  – Open Observatory of Network Interference (opt-in measurements of network connections)
  – Encore paper
Encore web-based measurements

- Burnett-Feamster study
- Embed in other non-blocked web pages measurement functionality of suspect censorship targets
- Use cross-site embeddings and browser side-channels
1. Origin serves page to client containing measurement task

2. Client renders page and executes measurement task

3. Task issues a cross-origin request for a resource on measurement target

4. Censor may filter request or response
Measure Web censorship

By adding a single line of code to your Web site, visitors of your site will automatically contribute data about how they experience Web censorship:

<iframe src="//encore.noise.gatech.edu/task.html" width="0" height="0" style="display: none"></iframe>

Learn more about Encore  Read the SIGCOMM 2015 paper  Encore settings
The results:

Confirmed blocking of:

- youtube.com in Pakistan, Iran, China
- Twitter.com, facebook.com in China, Iran
Statement from the SIGCOMM 2015 Program Committee: The SIGCOMM 2015 PC appreciated the technical contributions made in this paper, but found the paper controversial because some of the experiments the authors conducted raise ethical concerns. The controversy arose in large part because the networking research community does not yet have widely accepted guidelines or rules for the ethics of experiments that measure online censorship. In accordance with the published submission guidelines for SIGCOMM 2015, had the authors not engaged with their Institutional Review Boards (IRBs) or had their IRBs determined that their research was unethical, the PC would have rejected the paper without review. But the authors did engage with their IRBs, which did not flag the research as unethical. The PC hopes that discussion of the ethical concerns these experiments raise will advance the development of ethical guidelines in this area. It is the PC’s view that future guidelines should include as a core principle that researchers should not engage in experiments that subject users to an appreciable risk of substantial harm absent informed consent. The PC endorses neither the use of the experimental techniques this paper describes nor the experiments the authors conducted.
How would you avoid censorship?

- IP filtering
- DNS filtering / redirection
- URL filtering
- Packet filtering (search keywords in TCP packets)
- Protocol filtering (detect Tor protocol)
Preventing censorship

- End-to-end encryption (HTTPS, SSH)

- What does this protect?
- What does it leak?
Tor
(The Onion Router)

IP: 1.2.3.4

IP: 5.6.7.8

Tor Node 7.8.9.1

Tor Node 9.1.1.2

Tor Node 8.9.1.1

Censorship equipment

National internet

Other major backbone

Other major backbone
Onion routing: the basic idea

Tor implements more complex version of this basic idea
All data tunneled over point-to-point TLS connections
Tor
(The Onion Router)

National internet

Other major backbone

IP: 1.2.3.4

Tor Node
7.8.9.1

Tor Node
9.1.1.2

Tor Node
8.9.1.1

Other major backbone

IP: 5.6.7.8

Censorship equipment
Directly connecting users from the Islamic Republic of Iran

The Tor Project - https://metrics.torproject.org/
### Iran DPI blocking of Tor

- Tor point-to-point connections use TLS

#### National internet

**IP:** 1.2.3.4

**Censorship equipment**

#### Other major backbone

**IP:** 7.8.9.1

#### Tor Node

```
<table>
<thead>
<tr>
<th>No.</th>
<th>Time</th>
<th>Source</th>
<th>Destination</th>
<th>Protocol</th>
<th>Length</th>
<th>Info</th>
</tr>
</thead>
<tbody>
<tr>
<td>359</td>
<td>2.138821000</td>
<td>128.105.35.160</td>
<td>173.194.46.114</td>
<td>TLSv1.2</td>
<td>503</td>
<td>Client Hello</td>
</tr>
<tr>
<td>362</td>
<td>2.140902000</td>
<td>128.105.35.160</td>
<td>173.194.46.122</td>
<td>TLSv1.2</td>
<td>290</td>
<td>Client Hello</td>
</tr>
<tr>
<td>369</td>
<td>2.154594000</td>
<td>128.105.35.160</td>
<td>173.194.121.33</td>
<td>TLSv1.2</td>
<td>285</td>
<td>Client Hello</td>
</tr>
<tr>
<td>371</td>
<td>2.155001000</td>
<td>128.105.35.160</td>
<td>173.194.121.42</td>
<td>TLSv1.2</td>
<td>291</td>
<td>Client Hello</td>
</tr>
</tbody>
</table>
```

- Secure Sockets Layer
  - TLSv1.2 Record Layer: Handshake Protocol: Client Hello
    - Content Type: Handshake (22)
    - Version: TLS 1.0 (0x0301)
    - Length: 512
    - Handshake Protocol: Client Hello
      ```
      0040 63 8f 16 03 01 02 00 01 00 01 fc 03 03 1c 96 bf c ........
      0050 1a c0 ea b3 bc eb 04 45 d7 09 12 4d a1 6c 32 30 ........e ...M.120
      0060 6a 29 26 38 5e 65 07 a7 0a 72 ed e8 11 20 98 5f j?&e... r.... _
      0070 19 19 04 6f 38 2d 49 a1 00 a2 89 a9 4d 30 dd cc ...o6-I....M0..
      0080 e0 c8 3a 95 ab ed 76 29 ff 8b 0e e9 db 11 60 28 ........v) .......
      0090 c0 2b c0 2f 00 9e cc 14 cc 13 c0 0a c0 09 c0 13 .+/.... .......
      ```
TLS Handshake

Client

Pick random Nc

Check CERT using CA public verification key

Pick random PMS

C <- E(pk, PMS)

Bracket notation means contents encrypted

Server

ClientHello, MaxVer, Nc, Ciphers/CompMethods

ServerHello, Ver, Ns, SessionID, Cipher/CompMethod

CERT = (pk of server, signature over it)

C

ChangeCipherSpec,
{ Finished, PRF(MS, “Client finished” || H(transcript)) }

ChangeCipherSpec,
{ Finished, PRF(MS, “Server finished” || H(transcript’)) }

MS <- PRF(PMS, “master secret” || Nc || Ns )

PMS <- D(sk, C)
Iran DPI blocking of Tor

• Tor point-to-point connections use TLS
• Use DPI to filter Tor connections:
  – Tor certificates have short expiration date
  – Most websites have long expiration date
  – Shut down those connections with short expiration dates
  – https://blog.torproject.org/blog/update-internet-censorship-iran
• Tor fixed via longer expiration dates
• Later in 2012: blocking/degrading all TLS connections
Great Firewall targeting of Tor (circa 2011 and before)

• Enumerate Tor relays and filter them

![Number of relays graph]

Relay is publicly listed Tor node
Bridge is Tor node not publicly listed

The Tor Project - https://metrics.torproject.org/
Number of directory requests to directory mirror trusted

https://torproject.org
Great Firewall targeting of Tor (circa 2011 – today)

Admin noticed weird connections from China 2011

TLS connections with particular ciphersuites flagged for active probing

If remote server speaks Tor then add its IP address to blacklist

From [Winter, Lindskog 2012]

https://gist.github.com/da3c7a9af01d74cd7de7
TLS Handshake

Pick random Nc

Check CERT using CA public verification key

Pick random PMS

C <- E(pk, PMS)

Bracket notation means contents encrypted

ClientHello, MaxVer, Nc, Ciphers/CompMethods

ServerHello, Ver, Ns, SessionID, Cipher/CompMethod

CERT = (pk of server, signature over it)

C

ChangeCipherSpec,
{ Finished, PRF(MS, “Client finished” || H(transcript)) }

ChangeCipherSpec,
{ Finished, PRF(MS, “Server finished” || H(transcript’)) }

MS <- PRF(PMS, “master secret” || Nc || Ns )

PMS <- D(sk, C)
Great Firewall targeting of Tor (circa 2011 – today)

Admin noticed weird connections from China 2011

TLS connections with particular ciphersuites flagged for active probing

If remote server speaks Tor then add its IP address to blacklist

From [Winter, Lindskog 2012]

https://gist.github.com/da3c7a9af01d74cd7de7
How would you defeat DPI-based tool / protocol identification?

Make hard to detect as Tor:
- randomizers (obfsproxy)
- protocol mimicry
- tunneling

Mimicry & tunneling related to steganography
Great Firewall targeting of Tor (circa 2011 – today)

Ensafi et al. IMC 2015 follow-up study

• Active measurements, log file analysis, etc.
• China is checking obfsproxy3 bridges
• Hijack IP addresses to perform active probing
• DPI is stateful but does not reconstruct TCP streams
Censorship summary

• Nation-state censorship apparatuses are technologically sophisticated
• Bit of an arms race between circumvention community (activists, academics, USG)
• Lots of open research questions
  – Better steganography in face of DPI
  – Proxy distribution without enumeration
  – ...
Directly connecting users from Egypt
From BlueCoat:

• Our awareness of the presence of these ProxySG appliances in Syria came from reviewing online posts made by so-called “hacktivists” that contained logs of internet usage which appear to be generated by ProxySG appliances. We believe that these logs were obtained by hacking into one or more unsecured third-party servers where the log files were exported and stored. We have verified that the logs likely were generated by ProxySG appliances and that these appliances have IP addresses generally assigned to Syria. We do not know who is using the appliances or exactly how they are being used. We currently are conducting an internal review and also are working directly with appropriate government agencies to provide information on this unlawful diversion.
Directly connecting users from the Syrian Arab Republic

The Tor Project - https://metrics.torproject.org/
“Twitter, mwitter!”

Directly connecting users from Turkey

The Tor Project - https://metrics.torproject.org/
Protocol identification via deep-packet inspection (DPI)

DPI users want to identify protocol X

X = TLS or Tor then throttle connection

X = HTTP then leave it alone

X = ??? then throttle traffic

Check packet contents against regular expressions

/^\x16\x03[\x00\x01\x02]..\x02...\x03[\x00\x01\x02]|...? .*/

Free translation: Does packet include “I’m TLS 1.1” ?
Scenario:
DPI system only allows HTTP traffic unfettered

Obsfproxy (built into Tor): encrypt all bits sent over network (no plaintext bits)
  - Really fast
  - But DPI will flag traffic as ???

Want way to force DPI to classify traffic incorrectly as HTTP
So-called “misclassification attacks” against DPI
Surveying modern DPI systems

<table>
<thead>
<tr>
<th>System</th>
<th>Look at ports?</th>
<th>TCP stream reassembly</th>
<th>Uses regex’s</th>
<th>Use’s C/C++</th>
</tr>
</thead>
<tbody>
<tr>
<td>AppID</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>L7-filter</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Yaf</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Bro</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>nProbe</td>
<td>No</td>
<td>Yes</td>
<td>Not explicitly</td>
<td>Yes</td>
</tr>
<tr>
<td>Proprietary*</td>
<td>Yes</td>
<td>Yes</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

* Hint: it’s a serious product (~$10k) and similar ones seem to be used in Iran.

Can we build encryption schemes that fool regex-based systems?
How would you censor web requests?

- IP filtering
- DNS filtering / redirection
- URL filtering
- Packet filtering (search keywords in TCP packets)
- Protocol filtering (detect Tor protocol)
How would you censor web requests?

- IP filtering
- DNS filtering / redirection
- URL filtering
- Packet filtering (search keywords in TCP packets)
- Protocol filtering (detect Tor protocol)
Circumvention of filtering

- IP filtering
- Proxies
- DNS filtering / redirection
  - DNS proxy
- URL filtering or Packet filtering
  - Encryption / Tunneling / obfuscation
- Protocol filtering
  - Obfuscation techniques