Censored Planet: Measuring Internet Censorship Globally and Continuously

Roya Ensaafi
AIMS 2018
Measuring Internet Censorship Globally

PROBLEM:
- How can we detect whether pairs of hosts around the world can talk to each other?
Measuring Internet Censorship Globally

PROBLEM:
- How can we detect whether pairs of hosts around the world can talk to each other?

STATE OF THE ART:
- Deploy hardware or software at hosts (RIPE Atlas, OONI probe)
- Ask people on the ground, or use VPNs, or research networks (PlanetLab)

THREE KEY CHALLENGES:
Coverage, ethics, and continuity
These machines blindly follow Internet protocol rules such as TCP/IP.

How can we leverage standard protocol behaviors to detect whether two distant hosts can communicate?

140 million public live IPv4 addresses
Measuring Internet Censorship Globally... Remotely!

PROBLEM:
- How can we detect whether pairs of hosts around the world can talk to each other... from somewhere else in the world?

Impossible!
Spooky Scan uses TCP/IP side channels to detect whether a user and a site can communicate (and in which direction packets are blocked)

Goal: **Detect blocking from off-path**

* TCP Idle Scan Antirez, (Bugtraq 1998)
* Detecting Intentional Packet Drops on the Internet via TCP/IP Side Channels
  Roya Ensafi, Knockel, Alexander, and Crandall (PAM ’14)
* Idle Port Scanning and Non-interference Analysis of Network Protocol Stacks Using Model Checking
  Roya Ensafi, Park, Kapur, and Crandall (Usenix Security 2010)
**Augur** is a follow up system that uses the same TCP/IP side channels to detect blocking from off-path.

**Goal:** Scalable, ethical, and statistically robust system to continuously detect blocking.

*Augur: Internet-Wide Detection of Connectivity Disruption*

P. Pearce*, R. Ensafi*, F. Li, N. Feamster, V. Paxson

(* joint first authors)*
TCP Handshake:

- SYN [IP ID: X]
- SYN/ACK [IP ID: Y]
- ACK [IP ID: X+1]

Port status is open/closed

Port status is open
Spooky Scan Requirements

“User” (Reflector)
Must maintain a *global* value for IP ID

Site
Open port and retransmitting SYN-ACKs

Measurement Machine
Must be able to spoof packets
Spooky Scan

Measurement machine

Reflector

Reflector IP ID

Site
Spooky Scan

No direction blocked
Spooky Scan

No direction blocked

Measurement machine

1 SYN/ACK

2 RST [IP ID: 7000]

Reflector IP ID: 7000

Site
Spooky Scan

No direction blocked

Measurement machine

1 SYN/ACK

2 RST [IP ID: 7000]

3 Spoofed SYN [src: Reflector IP]

Reflector

Reflector IP ID: 7000

Site
Spooky Scan

No direction blocked

Measurement machine

Reflector

Reflector IP ID: 7000

SYN/ACK

RST [IP ID: 7000]

Spoofed SYN [src: Reflector IP]

SYN/ACK

Site
Spooky Scan

No direction blocked
Spooky Scan

No direction blocked

Reflector IP ID: 7000 7001 7002

Measurement machine

Spoofed SYN [src: Reflector IP]

Reflector

Site

SYN/ACK

RST [IP ID: 7000]

RST [IP ID: 7002]
Spooky Scan

No direction blocked
Spooky Scan

Site-to-Reflector
Blocked
Spooky Scan

Reflector-to-Site
Blocked

SYN/ACK
Measurement machine

SYN/ACK
RST [IP ID: 7000]

RST [IP ID: 7002]

Spoofed SYN [src: ClientIP]

Reflector IP ID: 7000 7001 7002

Site

Reflector-to-Site
Blocked
Spooky Scan

Reflector-to-Site Blocked
**Spooky Scan**

### Site-to-Reflector Blocked

\[ \Delta \text{IP ID1} = 1 \]
\[ \Delta \text{IP ID2} = 1 \]

### No Direction Blocked

\[ \Delta \text{IP ID1} = 2 \]
\[ \Delta \text{IP ID2} = 1 \]

### Reflector-to-Site Blocked

\[ \Delta \text{IP ID1} = 2 \]
\[ \Delta \text{IP ID2} = 2 \]
Coping with Reflector IP ID Noise

Amplifying the signal
Effect of sending $N$ spoofed SYNs:

<table>
<thead>
<tr>
<th>Site-to-Reflector Blocked</th>
<th>No Direction Blocked</th>
<th>Reflector-to-Site Blocked</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta \text{IP ID1} = (1 + \text{noise})$</td>
<td>$\Delta \text{IP ID1} = (1 + N + \text{noise})$</td>
<td>$\Delta \text{IP ID1} = (1 + N + \text{noise})$</td>
</tr>
<tr>
<td>$\Delta \text{IP ID2} = \text{noise}$</td>
<td>$\Delta \text{IP ID2} = \text{noise}$</td>
<td>$\Delta \text{IP ID2} = (1 + N + \text{noise})$</td>
</tr>
</tbody>
</table>
Coping with Reflector IP ID Noise

Amplifying the signal

Effect of sending $N$ spoofed SYNs:

- **Site-to-Reflector Blocked**
  - $\Delta IP ID1 = (1 + \text{noise})$
  - $\Delta IP ID2 = \text{noise}$

- **No Direction Blocked**
  - $\Delta IP ID1 = (1 + N + \text{noise})$
  - $\Delta IP ID2 = \text{noise}$

- **Reflector-to-Site Blocked**
  - $\Delta IP ID1 = (1 + N + \text{noise})$
  - $\Delta IP ID2 = (1 + N + \text{noise})$

Repeating the experiment

To eliminate the effects of packet loss, sudden bursts of packets, ...
Insight: Some measurements much noisier than others.
Augur for Continuous Scanning

**Insight:** Some measurements much noisier than others.

**Probing Methodology:**

Until we have high enough confidence (or up to):

- For first 4s, query IPID every sec
  - Send 10 spoofed SYNs
  - Query IPID
- Query IPID
Augur for Continuous Scanning

**Insight:** Some measurements much noisier than others.

**Probing Methodology:**

Until we have high enough confidence (or up to):

- For first 4s, query IPID every sec
  - Send 10 spoofed SYNs
  - Query IPID
- Query IPID

**Repeat runs and use Seq. Hypothesis Testing to gradually build confidence.**
Defining a random variable:

\[ Y_n(S_i, R_j) = \begin{cases} 
1 & \text{if no IPID acceleration occurs} \\
0 & \text{if IPID acceleration occurs} 
\end{cases} \]
Augur: Sequential Hypothesis Testing

Defining a random variable:

\[ Y_n(S_i, R_j) = \begin{cases} 1 & \text{if no IPID acceleration occurs} \\ 0 & \text{if IPID acceleration occurs} \end{cases} \]

Calculate known outcome probabilities (priors):

Prior 1: Prob. of no IPID acceleration when there is blocking
Prior 2: Prob. of IPID acceleration when there is no blocking
Augur: Sequential Hypothesis Testing

Based on $\Lambda(Y)$, can we decide the blocking case?

- Site-to-Ref blocking
- No Blocking
- Ref-to-Site blocking

Maximum Likelihood Ratio

$$\Lambda(Y) = \prod_{n=1}^{N} \frac{Pr[Y_n|Blocking]}{Pr[Y_n|No\ Blocking]}$$
Augur Framework
Augur Framework

User input

- All responsive IPs
- Target countries

Reflector selection

Reflector Characterization
Augur Framework

- User input:
  - All responsive IPs
  - Target countries
  - Site address

- Processes:
  - Reflector selection
  - Reflector Characterization
  - Site characterization
Augur Framework

User input
- Target countries
- Site address

All responsive IPs

Reflector selection

Reflector Characterization

Probing

Scheduler

Site characterization
Augur Framework

- User input:
  - All responsive IPs
  - Target countries
  - Site address

- System output:
  - Ref-to-Site blocking
  - Site-to-Ref blocking
  - No blocking
  - Error

- Detection/Validation
  - Reflector selection
  - Reflector Characterization
  - Site characterization
  - Scheduler
Challenge: Need global vantage points from which to measure

Coverage

Scanning IPv4 on port 80:

- 22.7 million potential reflectors!

Compare: 10,000 in prior work (RIPE Atlas)
**Challenge:** Probing banned sites from users’ machines creates risk
**Challenge**: Probing banned sites from users’ machines creates risk

**Ethics**: Use only **infrastructure devices** to source probes

**THREE KEY CHALLENGES**: Coverage, ethics, and continuity

- Global IP ID: 22.7 million, 236 countries (and dependent territories)
- Two hops back from end user: 53,000, 180 countries
Augur doesn’t depend on end users’ availability, and routers have less downtime, allowing us to collect measurements continuously.

**Challenge:** Need to repeat measurements over time

**Continuity**

**TWO KEY CHALLENGES:** Coverage, ethics, and continuity
Running Augur In the Wild

Reflectors: 2,050
Sites: 2,134 (Citizen Lab list + Alexa Top-10K)
    Mix of sensitive and popular sites
Duration: 17 days
Measurements per reflector-site: 47
Overall # of measurements: 207.6 million
## Site-to-Reflector blocking

<table>
<thead>
<tr>
<th>No.</th>
<th>Site</th>
<th>% Refs</th>
<th>% Cnt.</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>hrcr.org</td>
<td>41.7</td>
<td>83.0</td>
<td>Human Rights</td>
</tr>
<tr>
<td>2.</td>
<td>alstrangers.[LJ].com</td>
<td>37.9</td>
<td>78.8</td>
<td>Militants</td>
</tr>
<tr>
<td>3.</td>
<td>varlamov.ru</td>
<td>37.7</td>
<td>78.0</td>
<td>Foreign relations</td>
</tr>
<tr>
<td></td>
<td>nordrus-norna.[LJ].com</td>
<td></td>
<td></td>
<td>Hate speech</td>
</tr>
<tr>
<td>4.</td>
<td><a href="http://www.stratcom.mil">www.stratcom.mil</a></td>
<td>37.5</td>
<td>78.6</td>
<td>Foreign relations</td>
</tr>
<tr>
<td>5.</td>
<td><a href="http://www.demonoid.me">www.demonoid.me</a></td>
<td>21.7</td>
<td>58.5</td>
<td>P2P file sharing</td>
</tr>
<tr>
<td>6.</td>
<td>amateurpages.com</td>
<td>21.2</td>
<td>57.9</td>
<td>Adult contents</td>
</tr>
<tr>
<td></td>
<td>voice.yahoo.jajah.com</td>
<td></td>
<td></td>
<td>Voice over IP</td>
</tr>
<tr>
<td></td>
<td>amtrak.com</td>
<td></td>
<td></td>
<td>ALEXA</td>
</tr>
</tbody>
</table>

### Interesting example:
- **amtrak.com** was blocked for 21% of reflectors, 57% of countries (ranked 6) → Collateral damage
Top Blocked Sites

Reflector-to-site Blocked

Interesting example:
- nsa.gov was blocked for 7.4% of reflectors, 23% of countries (ranked 1)

Note: Some servers discriminate by providing their services to specific regions

Examples: Dating sites, banking sites, or sites that have to follow embargo rules

### Reflector-to-site blocking

<table>
<thead>
<tr>
<th>No.</th>
<th>Site</th>
<th>% Refs</th>
<th>% Cnt.</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>nsa.gov</td>
<td>7.4</td>
<td>23.3</td>
<td>US Gov.</td>
</tr>
<tr>
<td>2.</td>
<td>scientology.org</td>
<td>2.2</td>
<td>6.9</td>
<td>Minority faiths</td>
</tr>
<tr>
<td>3.</td>
<td>goarch.org</td>
<td>1.9</td>
<td>4.4</td>
<td>Minority faiths</td>
</tr>
<tr>
<td>4.</td>
<td>yandex.ru</td>
<td>1.8</td>
<td>3.8</td>
<td>Freedom of Expression</td>
</tr>
<tr>
<td>5.</td>
<td>hushmail.com</td>
<td>1.8</td>
<td>4.4</td>
<td>Free email</td>
</tr>
<tr>
<td>6.</td>
<td>carnegieendowment.org</td>
<td>1.6</td>
<td>4.4</td>
<td>Political reforms</td>
</tr>
</tbody>
</table>
Augur is a system that uses TCP/IP side channels to continuously detect blocking.

- Reduce risks by using only infrastructure devices to source probes
- Can use more than 53,000 to cover more than 180 countries
Side Channels at Other Network Layers

Network interference happens at all layers

What’s new on cnn.com?

Resolver

DNS A query for cnn.com

HTTP requests

(opt) TLS handshake

TCP handshake

IP routing
**Satellite** is a system that uses DNS open resolvers to detect whether a user can resolve a domain accurately

**Goal:** Scalable, ethical, and statistically robust system to continuously detect DNS level manipulation

* Global Measurement of DNS Manipulation, Pearce, Jones, Li, Ensafi, Feamster, Paxson, USENIX Security, August 2017
Deploying Satellite

**Challenge:**
Identify “wrong” DNS responses

**Coverage:**
- Scan IPv4 for open resolvers: 4.2 M, 232 countries

**Ethical:**
- Using resolvers reasonably attributed to Internet naming infrastructures: ~ 7k

**Continuity:**
- Satellite doesn’t depend on end users’ availability, and resolvers have less downtime

**Detecting DNS manipulation:**
- Using consistency and independent verifiability heuristics.
Network interference happens at all layers

What's new on cnn.com?

Resolver

DNS query for cnn.com

HTTP requests

(opt) TLS handshake

TCP handshake

IP routing
Network interference happens at all layers

Side Channels at Other Network Layers

Network interference happens at all layers

What's new on cnn.com?

DNS query for cnn.com

Resolver

HTTP requests

(opt) TLS handshake

TCP handshake

IP routing
Censored Planet, a system that provides a continual and global view of Internet censorship

- **Daily reachability measurements** for key websites from countries worldwide

- Data collected with Augur, Satellite, and Quack combined with **side channels at other network layers**

- Tools for mapping and **comparative analyses** across locations and time
Censored Planet: 
Measuring Internet Censorship
Globally and Continuously

Roya Ensafi
CAIDA, 2018