Measuring Censorship Everywhere
All the Time

Jeffrey Knockel
Jedidiah R. Crandall

Department of Computer Science
University of New Mexico
Goal

- Measure censorship everywhere all the time
- Problems:
  - no vantage points in country
  - not in right city/institution/building/etc.
Our solution

- Side channels: turn ordinary machines into vantage points
- Measure IP censorship off-path
- No software on server, client, or anywhere in between

1.2.3.4  ??  5.6.7.8
Previous Layer 4 (TCP) technique

Client

- Find client with *globally incrementing IP ID*

IP Header

<table>
<thead>
<tr>
<th>Version / IHL / TOS</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ID</strong></td>
<td>Flags / Fragment Offset</td>
</tr>
<tr>
<td>TTL / Protocol</td>
<td>Check</td>
</tr>
<tr>
<td>Source IP</td>
<td></td>
</tr>
<tr>
<td>Destination IP</td>
<td></td>
</tr>
</tbody>
</table>

- Windows XP, FreeBSD, etc. globally reuse this ID

Client → Server censored (+>1)

Server: SYNACK → Client

SYNACK → RST

RST → SYNACK

...
Layer 3 (IP) techniques

- Layer 3 already has enough side channels
- More general assumptions
- One technique we have is for Linux servers
- Question: can some address talk to some Linux server?
- If that address responds to pings, then we can measure this!
IP fragments

• Utilize Linux's *fragment cache* behavior
• IP datagrams are split into fragments when they are too large to go over a medium
Fragment cache

- Fragments are kept in a cache until all fragments arrive and the datagram is complete.
- Linux has a “maximum distance” rule:
  - If I receive a fragment for datagram $d$ from address $X$.
  - Then I receive another 64 fragments from $X$.
  - If $d$ hasn't been completed yet, then its other fragments ain't ever coming.
- Bookkeeping! Side channel!
By way of example

Linux machine

Can P talk to L?

Pingable address

Albuquerque
Prime L

Linux machine L

Pingable address P

Received 63 fragments from P since

Spoof 63 fragment first-halves from P

Albuquerque
Spoof echo request

Linux machine

L

Pingable address

P

Received 63 fragments from P since

Albuquerque

Spoof large, fragmented echo request from L
Case: Censorship

Linux machine  
L

Received 63 fragments from P since

Pingable address  
P

Albuquerque
Case: No censorship

Linux machine
L

Fragmented echo reply

Received 65+ fragments from P since

Pingable address
P

Albuquerque
Complete datagrams

Linux machine L

Pingable address P

Spoof 63 fragment second-halfes from P (in same order)

Albuquerque
Censorship cases

- In censorship case:
  - Second halves complete datagram

- In no censorship case:
  - Second halves are too late!
  - The first halves are already gone
  - The second halves create new entries
To actually determine censorship

- Are those 63 entries in there or not?
- How much room is left?
- Send our own pings:
  - Room for (e.g.) 263 ⇒ Censorship
  - Room for (e.g.) 200 ⇒ No censorship
Deploying vantage points

- Almost 10% of IPv4 address space responds to large pings
  - Over 16% of China's space
- To deploy 10 vantage points...
  - Ping 100 random IP addresses
  - Which 10 respond to large pings?
  - That's it!
Ethical considerations

- Vantage points do not send pings—they respond to pings
- Measure an entire (e.g.) /24
  - Make it look like someone is ping sweeping