RUHR-UNIVERSITÄT BOCHUM

ON THE CHALLENGES OF GEOGRAPHICAL AVOIDANCE FOR TOR

Katharina Kohls
Kai Jansen, David Rupprecht, Thorsten Holz
Ruhr University Bochum

Christina Pöpper
NYU Abu Dhabi
Tor Anonymity System

Separate endpoints → Anonymity

230.125.195.89

www.sensitive-site.org
Traffic Analysis Attacks

Metadata leaks information → De-Anonymization

230.125.195.89
Direct Traffic Obfuscation

- Direct defenses are \textbf{expensive}:
  - Delay transmissions
  - Consume resources

\textbf{Mixing}: Additional Delays

\textbf{Cover Traffic}: Exhaust bandwidth
Alternatives

• Direct defenses are **expensive**:  
  • Delay transmissions  
  • Consume resources

**Mixing**: Additional Delays

**Cover Traffic**: Exhaust bandwidth

Are there alternative defenses?
Geographical Avoidance

The general concept.
General Concept

The server is here

You are here

The untrusted area is here
Standard Circuit

Server

Entry

Middle

Client

Exit
1. Detect connection through untrusted area
Use Better Circuit

1. Detect connection through untrusted area
2. Discard and create new circuit
How can we do this?

1. Detect connection through untrusted area
2. Discard and create new circuit
Timing Decisions

- Detect connection through untrusted area
  - Relays: GeoIP location data
  - Routing: Not transparent
  - Measure end-to-end timing

2. Z. Li, S. Herwig, and D. Levin, “DeTor: Provably Avoiding Geographic Regions in Tor,” in USENIX Security Symposium, USENIX’17
Estimate Worst Case

1. Find closest point in untrusted area
2. Measure distance between client and point
3. Assume speed, e.g., $\frac{2}{3}$ speed of light
4. Estimate RTT
Timing Decision

1. Find closest point in untrusted area
2. Measure distance between client and point
3. Assume speed, e.g., $\frac{2}{3}$ speed of light
4. Estimate RTT

- Use threshold for decisions
  - $R_{e2e} < R_{est}$  ✔
Timing Decision

1. Find closest point in untrusted area
2. Measure distance between client and point
3. Assume speed, e.g., $\frac{2}{3}$ speed of light
4. Estimate RTT

- Use threshold for decisions
  - $R_{e2e} < R_{est}$ ✓
  - $R_{e2e} \geq R_{est}$ ×
Challenges of Geo Avoidance
Considerations for the system design.
Three Classes of Challenges

1. **Network Diversity**
   1. *Distribution of Relays*
   2. Varying Connections Lengths
   3. Connection Failures

2. **Ground Truth**
   1. GeoIP Location Errors
   2. Assymetric Routes
   3. Intransparent Transmission Characteristics

3. **Deployment**
   1. Maintaining Tor's Performance and Security
   2. Using Reliable Information Sources

*On the Challenges of Geographical Avoidance for Tor*
Three Classes of Challenges

1. Network Diversity
   1. Distribution of Relays
   2. *Varying Connections Lengths*
   3. Connection Failures

2. Ground Truth
   1. GeoIP Location Errors
   2. Asymmetric Routes
   3. Intransparent Transmission Characteristics

3. Deployment
   1. Maintaining Tor's Performance and Security
   2. Using Reliable Information Sources

---

On the Challenges of Geographical Avoidance for Tor
Three Classes of Challenges

1. **Network Diversity**
   1. Distribution of Relays
   2. Varying Connections Lengths
   3. *Connection Failures*

2. **Ground Truth**
   1. GeoIP Location Errors
   2. Assymetric Routes
   3. Intransparent Transmission Characteristics

3. **Deployment**
   1. Maintaining Tor's Performance and Security
   2. Using Reliable Information Sources

~11% of circuits failed
12,500 out of 105,889
Three Classes of Challenges

1. **Network Diversity**
   1. Distribution of Relays
   2. Varying Connections Lengths
   3. Connection Failures

2. **Ground Truth**
   1. *GeoIP Location Errors*
   2. Asymmetric Routes
   3. Intransparent Transmission Characteristics

3. **Deployment**
   1. Maintaining Tor’s Performance and Security
   2. Using Reliable Information Sources

---

**GeoIP:**
- 6% (330/6042): provably false location
- 3.2% (194/6042): updated country code
Designing the Avoidance System

Prototype

Deployability

Simulation Model

Comparison of Approaches

Network Diversity

Ground Truth

Deployment

On the Challenges of Geographical Avoidance for Tor
Prototype: TrilateraTor
Considering the challenges.
Considering the Challenges

Prototype

- Network Diversity
- Ground Truth
- Deployment

Empirical Timing Decisions

1

Verification of Relay Locations

2

Secure Information Sources

3
Network Diversity: Timing Decisions

Upper Bound Decision

Distance: 4,384 km
2,724 miles

Speed: 0.66c (speed of light)

Time: 14.62 ms
Empirical Timing Decisions

**Upper Bound Decision**

Distance:
4,384 km
2,724 miles

Speed:
0.66c (speed of light)

Time:
14.62 ms

**TrilateraTor**

Time:
Measure circuits from remote servers
Hop Relations Table

Upper Bound Decision

Distance: 4,384 km
2,724 miles

Speed: 0.66c (speed of light)

Time: 14.62 ms

TrilateraTor

Timing Relations

<table>
<thead>
<tr>
<th>From</th>
<th>To</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>13</td>
</tr>
<tr>
<td>A</td>
<td>C</td>
<td>9</td>
</tr>
<tr>
<td>B</td>
<td>C</td>
<td>12</td>
</tr>
<tr>
<td>A</td>
<td>D</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

On the Challenges of Geographical Avoidance for Tor
Considering the Challenges

Prototype

Network Diversity

Ground Truth

Deployment

Empirical Timing Decisions

Verification of Relay Locations

Secure Information Sources
Ground Truth: Relay Locations

- Measuring relay positions
  - Send ICMP probes to relays
  - Use multiple reference points
  - Estimate position using trilateration
Considering the Challenges

Prototype

Network Diversity

Ground Truth

Deployment

Empirical Timing Decisions

Verification of Relay Locations

Secure Information Sources
Deployment: Timing Measurements

- Prior work: Probe the entire circuit
- Circuit is not checked at this point
- Two major issues:
  - Security: Reveals endpoint to adversary
  - Performance: Requires additional measurements

2. Z. Li, S. Herwig, and D. Levin, "DeTor: Provably Avoiding Geographic Regions in Tor," in USENIX Security Symposium, USENIX’17
Alternative: Handshake Timings

On the Challenges of Geographical Avoidance for Tor
Secure Information Sources

- No additional measurements
- Delivers end-to-end timing of circuit
- Does not reveal connection endpoint
Technical Concept

Network Side

Client Side
Two Types of Measurements

Network Side

Distributed Measurements

Client Side

Circuit Evaluation

Circuit Measurement
Decision Data

Network Side

Distributed Measurements

Relay Verification (ICMP)
Empirical Estimates (TCP)

Client Side

Circuit Evaluation

Decision

Circuit Measurement

End-to-End Timing of Circuit

\[ R_{est} \quad \overset{\cdots}{\longrightarrow} \quad R_{e2e} \quad < \quad R_{est} \quad ? \quad \overset{\cdots}{\longleftarrow} \quad R_{e2e} \]
Experiments
Gathering empirical data, comparing approaches.
Metrics: How to measure what we achieved

1. Restrictive avoidance decisions harm the network.
2. Static thresholds are not realistic.

- We measure:
  - What if…? Loss of bandwidth and circuits in different scenarios.
  - Time Ratio: Difference between the measured and the estimated time.
What if…?
What if Germany was forbidden area?

Empirical Decision

- Rejected Circuits: 71%
- Rejected Bandwidth: 74%
What if Germany was forbidden area?

**Empirical Decision**
- Rejected Circuits: 71%
- Rejected Bandwidth: 74%

**Static Decision**
- Rejected Circuits: 90%
- Rejected Bandwidth: 86%
Limit Performance Impairments

Remaining Resources

Germany

Average

22%

$\rightarrow 216 \text{ MBit/s bandwidth saved}$
Conclusion

Lessons learned.
Challenges of Geographical Avoidance

3 Classes of Challenges

1. Network Diversity
2. Ground Truth
3. Deployment
Designing an Avoidance System

3 Classes of Challenges
1. Network Diversity
2. Ground Truth
3. Deployment

Main Features
1. Empirical Decisions
2. Verification of Locations
3. Secure Information Sources

On the Challenges of Geographical Avoidance for Tor
Prototype with Tradeoff

3 Classes of Challenges
1. Network Diversity
2. Ground Truth
3. Deployment

Main Features
1. Empirical Decisions
2. Verification of Locations
3. Secure Information Sources

Evaluation
1. Time Ratio for Decision Tradeoff
2. What-if Analysis

On the Challenges of Geographical Avoidance for Tor
Thank You! Questions?

3 Classes of Challenges
1. Network Diversity
2. Ground Truth
3. Deployment

Main Features
1. Empirical Decisions
2. Verification of Locations
3. Secure Information Sources

Evaluation
1. Time Ratio for Decision Tradeoff
2. What-if Analysis

On the Challenges of Geographical Avoidance for Tor
Appendix

More information
1. Network Diversity: Connection Lengths

EU ↔ EU
4,384 km
2,724 miles
1. Network Diversity: Connection Lengths

NA ↔ NA
19,210 km
11,937 miles
1. Network Diversity: Connection Failures

~11% of circuits failed
12,500 out of 105,889
Verification of Relay Locations

- Measuring relay positions
  - Send ICMP probes to relays
  - Use multiple reference points
  - Estimate position using trilateration

**Problem:**
Which position is more precise?
Physical Proof

- Measuring relay positions
  - Send ICMP probes to relays
  - Use multiple reference points
  - Estimate position using trilateration

**Speed of light proof**
1. Measure RTT from server to relay
2. Compute upper bound threshold with $c$
   1. Measured Speed ≤ Speed of light? ✓
   2. Measured Speed > Speed of light? ✗
3. Violation: Update GeoIP location with estimate

S. Capkun and J. P. Hubaux, “Secure Positioning of Wireless Devices with Application to Sensor Networks,” in Annual Joint Conference of the IEEE Computer and Communications Societies
Comparison of Approaches

![Comparison of Approaches Graph]

On the Challenges of Geographical Avoidance for Tor
Prototype Simulation

On the Challenges of Geographical Avoidance for Tor
Time Ratio

\[ \frac{R_{est}}{R_{e2e}} \]

On the Challenges of Geographical Avoidance for Tor
Decision Threshold

\[
\frac{R_{est}}{R_{e2e}}
\]

On the Challenges of Geographical Avoidance for Tor
Handshake Overhead

On the Challenges of Geographical Avoidance for Tor
Measurement Statistics

Stability of Results

<table>
<thead>
<tr>
<th>Type</th>
<th>Iteration</th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
<th>Duration</th>
<th>#Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCP</td>
<td>1</td>
<td>287</td>
<td>288</td>
<td>158</td>
<td>5 days</td>
<td>223,070</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>359</td>
<td>335</td>
<td>180</td>
<td>7 days</td>
<td>134,370</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>327</td>
<td>295</td>
<td>185</td>
<td>8 days</td>
<td>275,509</td>
</tr>
<tr>
<td>ICMP</td>
<td>1</td>
<td>99</td>
<td>67</td>
<td>98</td>
<td>1 day</td>
<td>27,274</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>56</td>
<td>18</td>
<td>77</td>
<td>1 day</td>
<td>62,643</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>136</td>
<td>128</td>
<td>102</td>
<td>2 days</td>
<td>1,837,761</td>
</tr>
</tbody>
</table>

Measurement Overhead

- Approx. 2.8 Mio. daily Tor users, 121.5 Gbit/s average consumed bandwidth
- TrilateraTor consumes $6.24 \times 10^{-7}\%$ of daily bandwidth and $4 \times 10^{-4}\%$ of circuits
Experimental Setup

- 8 Server instances
- **Hop Estimates** $R_{e2e}$: 16,500 relay combinations
  - 1,945 Entries, 3,724 Middles, 893 Exits
- **Circuit RTT** $R_{est}$: 70,081 circuits, 275,509 measurements
  - 1,670 Entries, 2,712 Middles, 735 Exits (artificial circuits)
  - 135,924 reference circuits