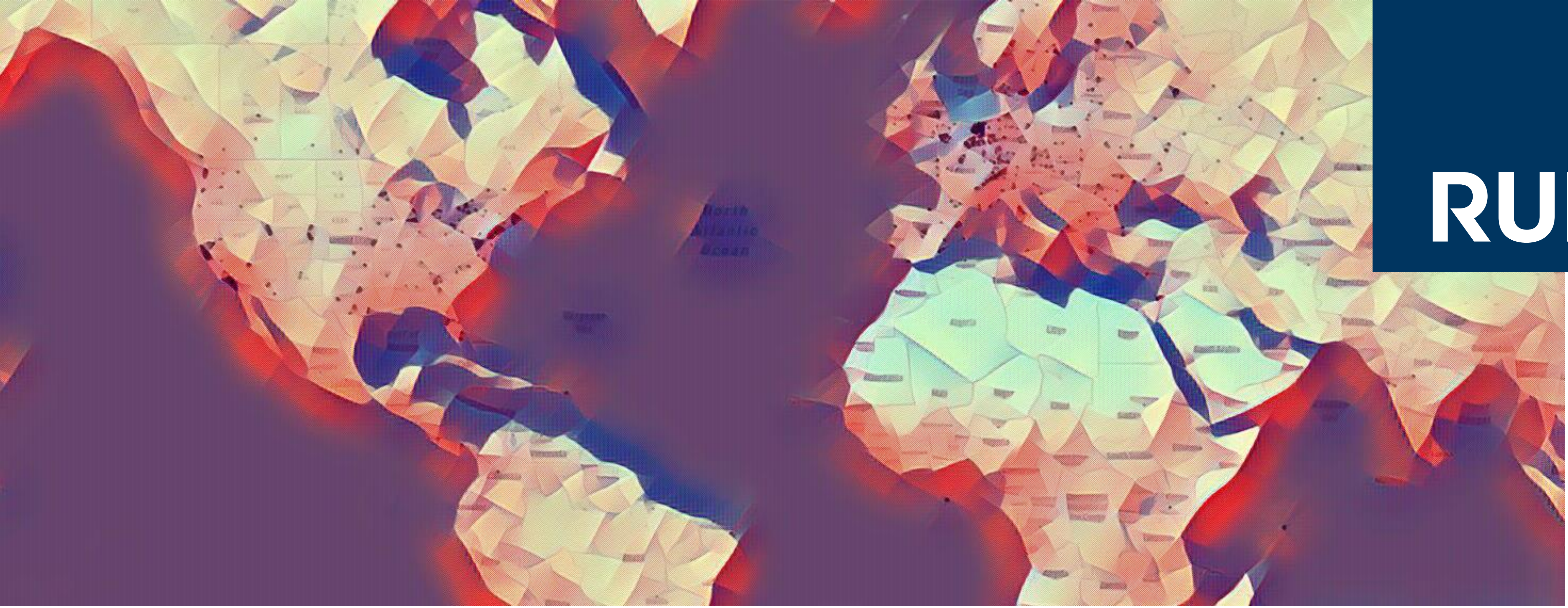


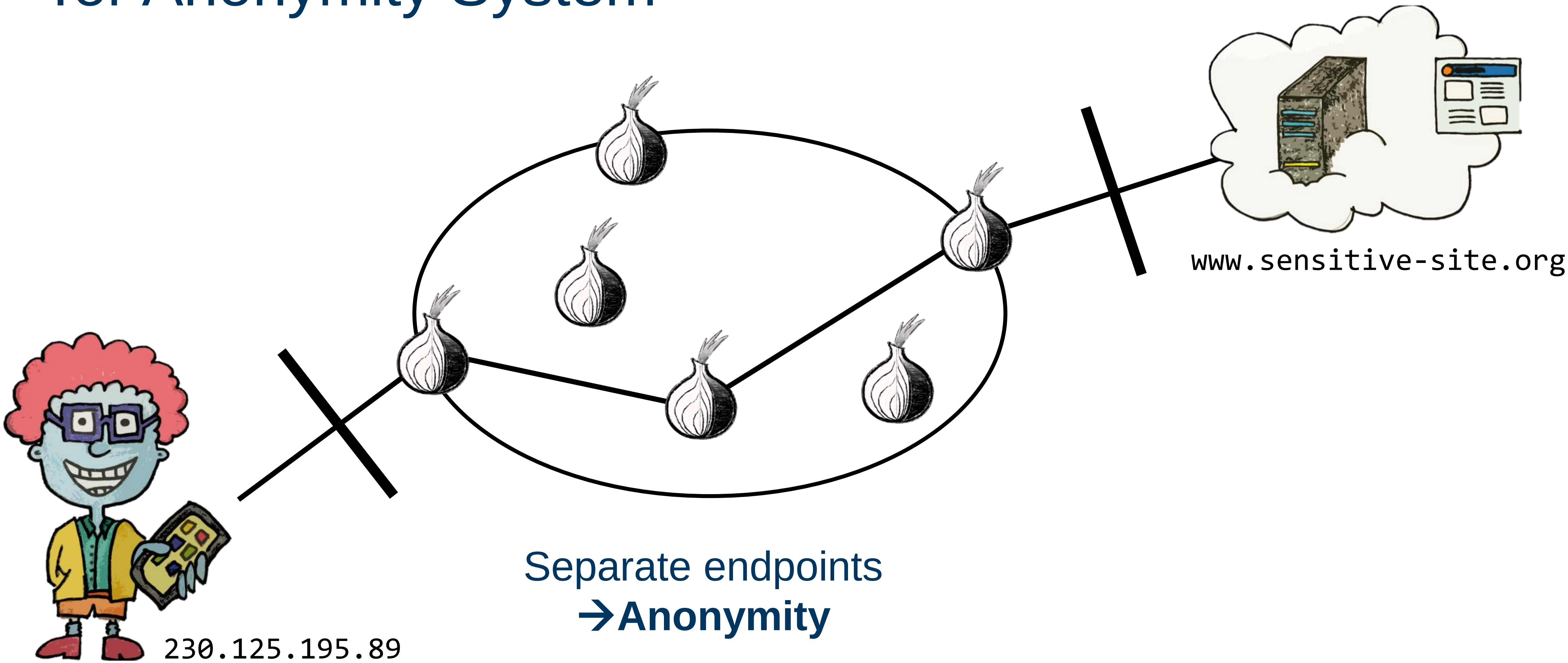
RUHR-UNIVERSITÄT BOCHUM

ON THE CHALLENGES OF GEOGRAPHICAL AVOIDANCE FOR TOR

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

Christina Pöpper
NYU Abu Dhabi

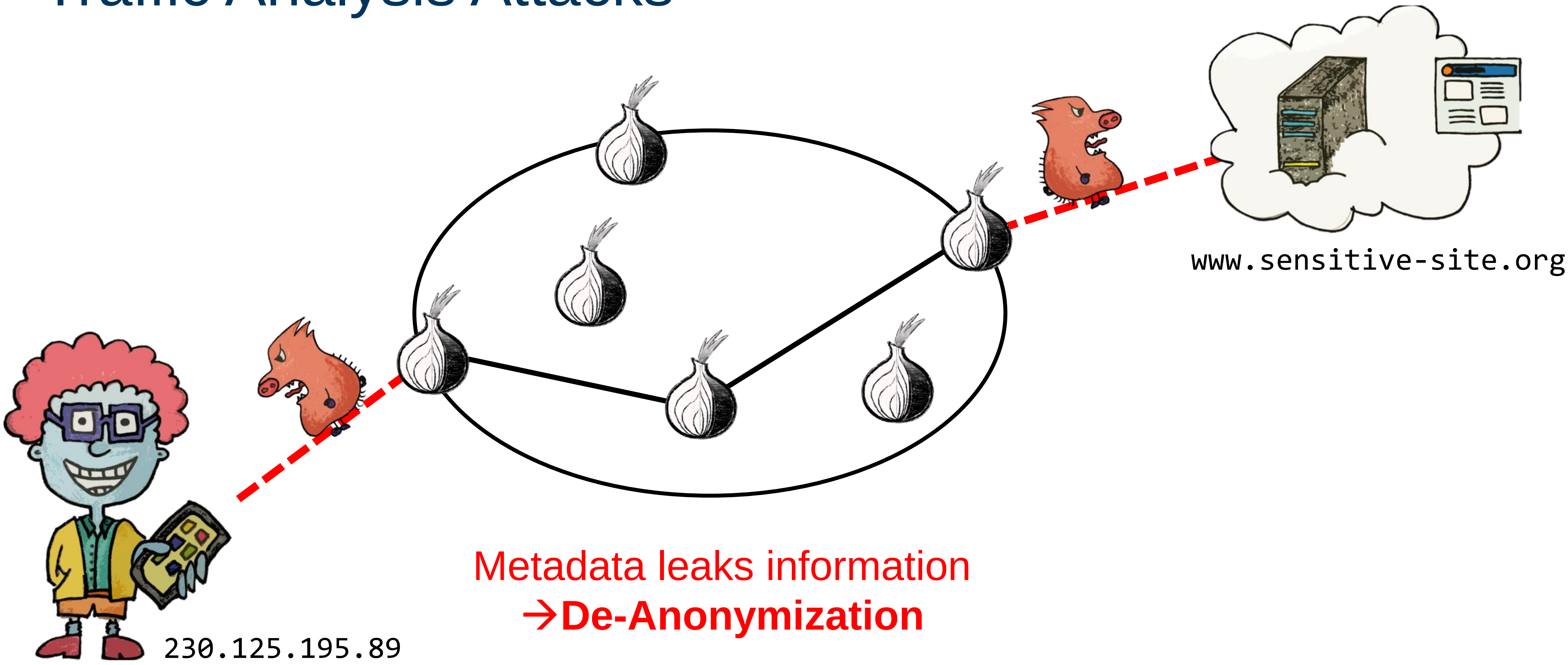
Tor Anonymity System



Separate endpoints

→**Anonymity**

Traffic Analysis Attacks



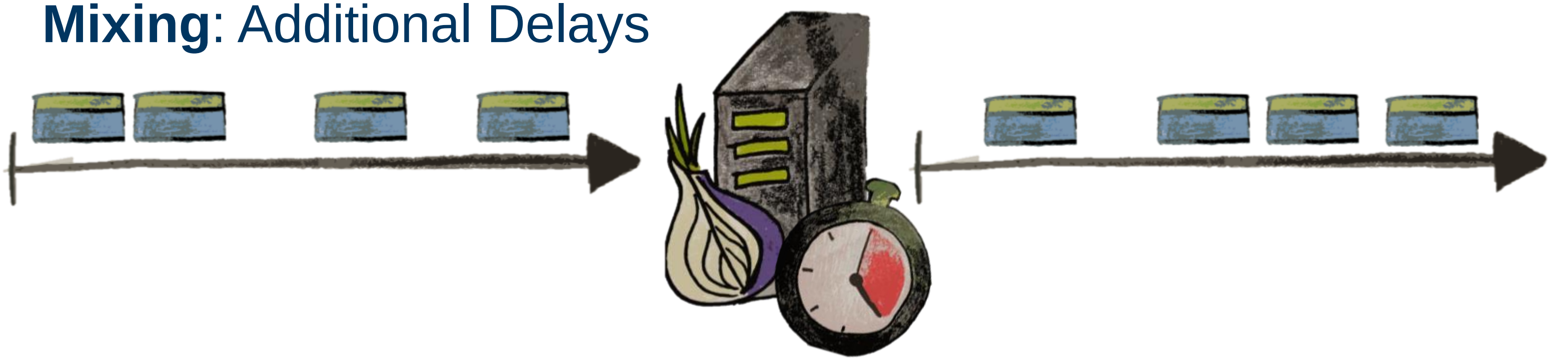
Metadata leaks information

→**De-Anonymization**

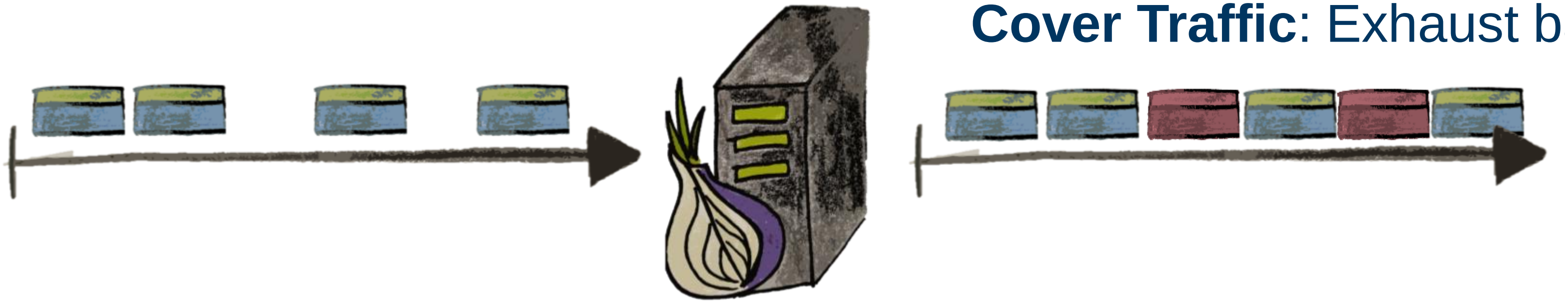
Direct Traffic Obfuscation

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

Mixing: Additional Delays



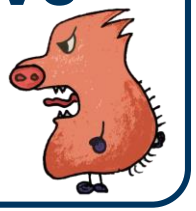
Cover Traffic: Exhaust bandwidth



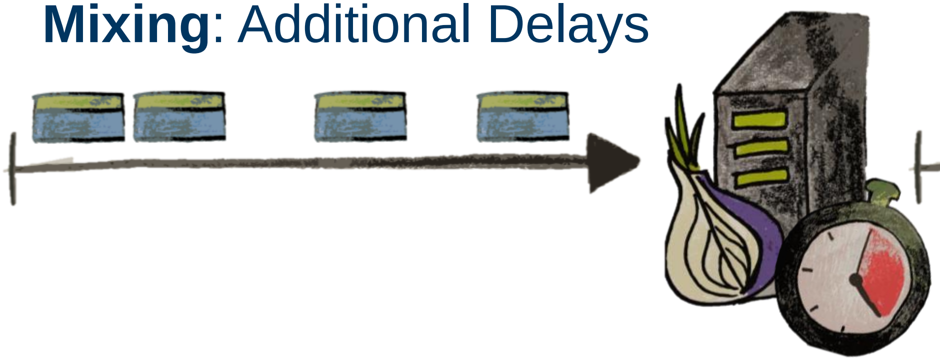
Alternatives

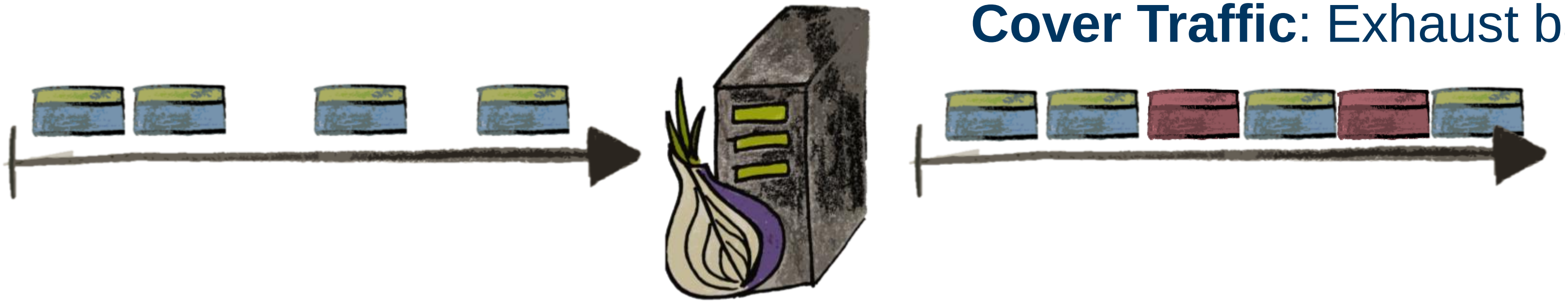
- Direct defenses are **expensive:**
  - Delay transmissions
  - Consume resources

Are there alternative defenses?



Mixing: Additional Delays



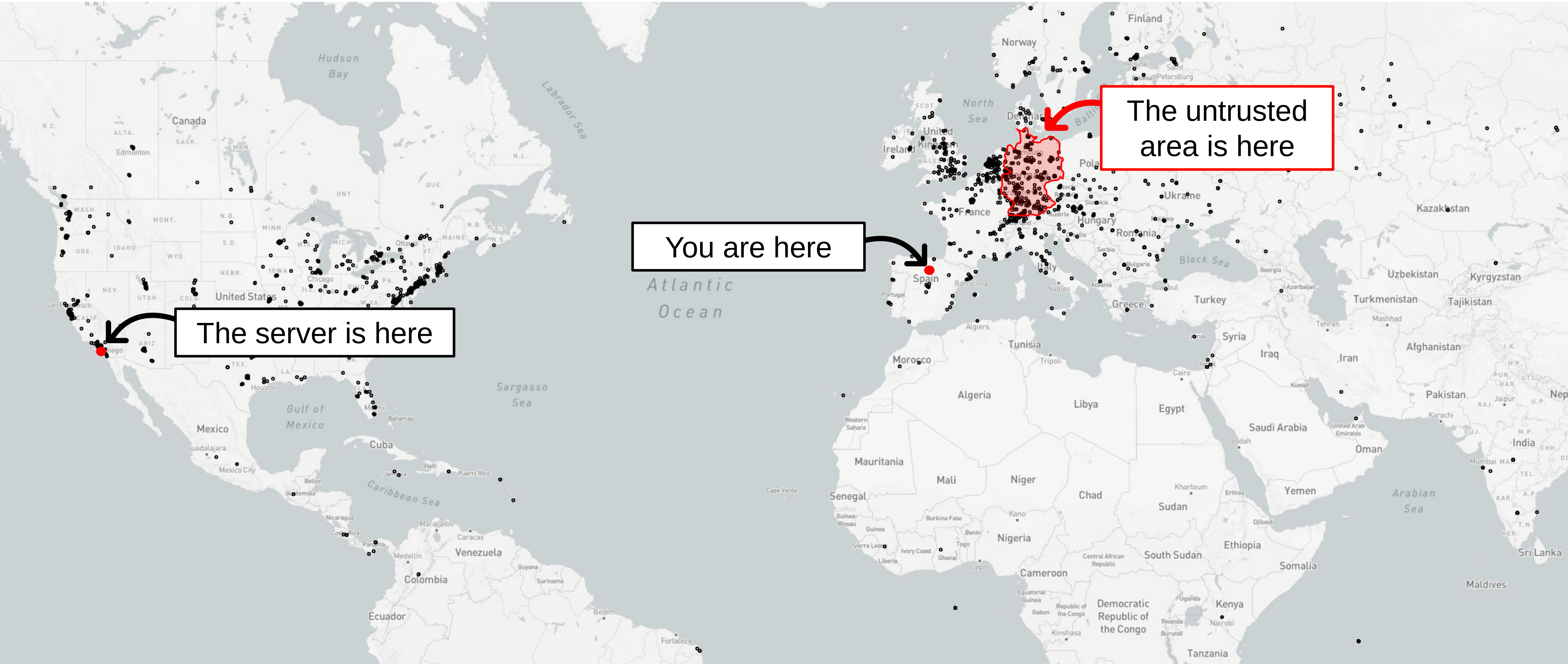


Cover Traffic: Exhaust bandwidth

Geographical Avoidance

The general concept.

General Concept

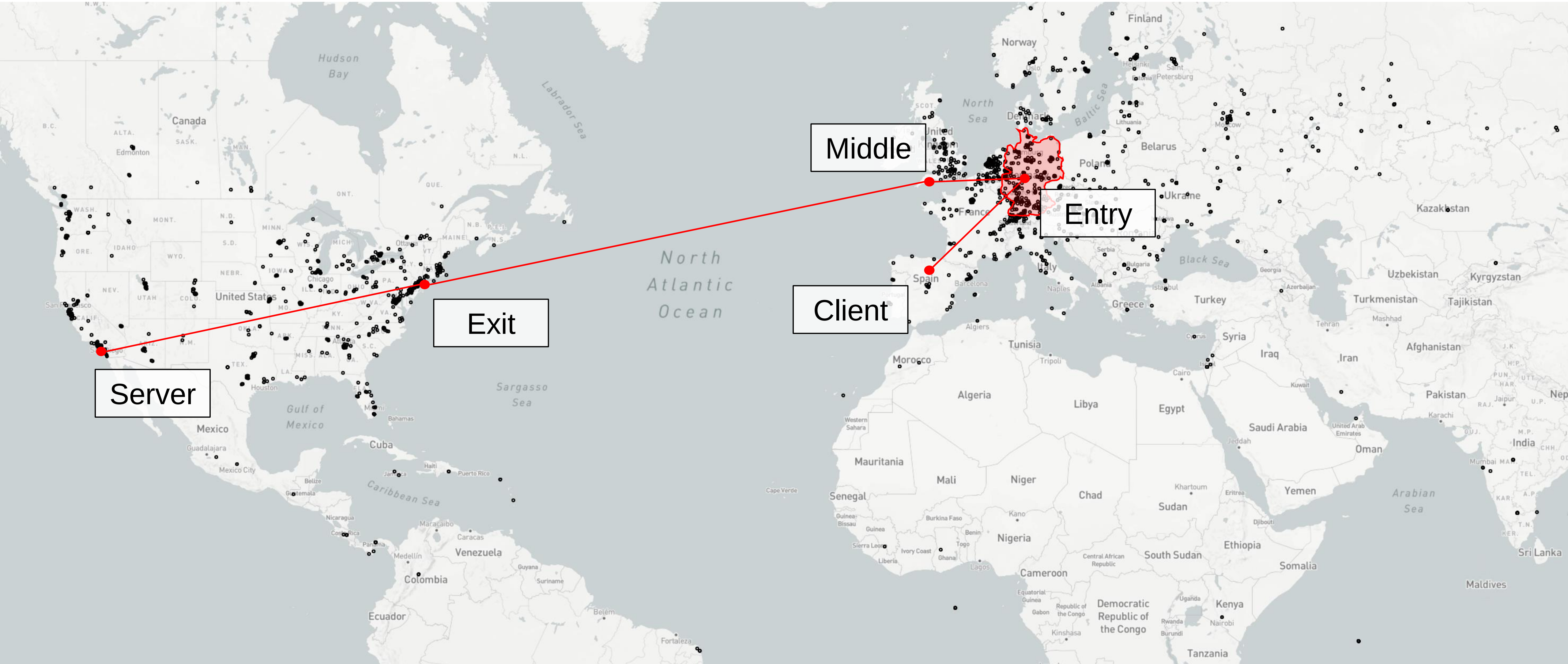


The untrusted area is here

You are here

The server is here

Standard Circuit



Middle

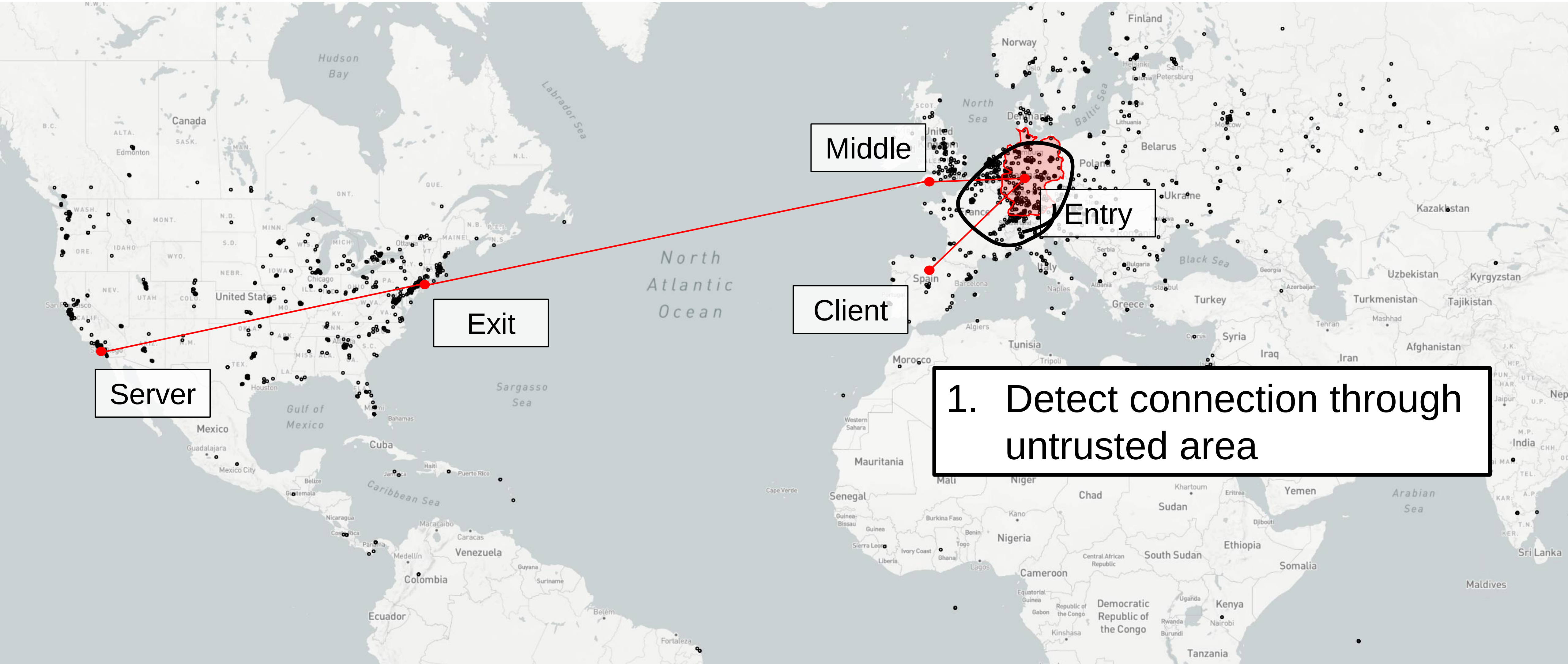
Entry

Client

Exit

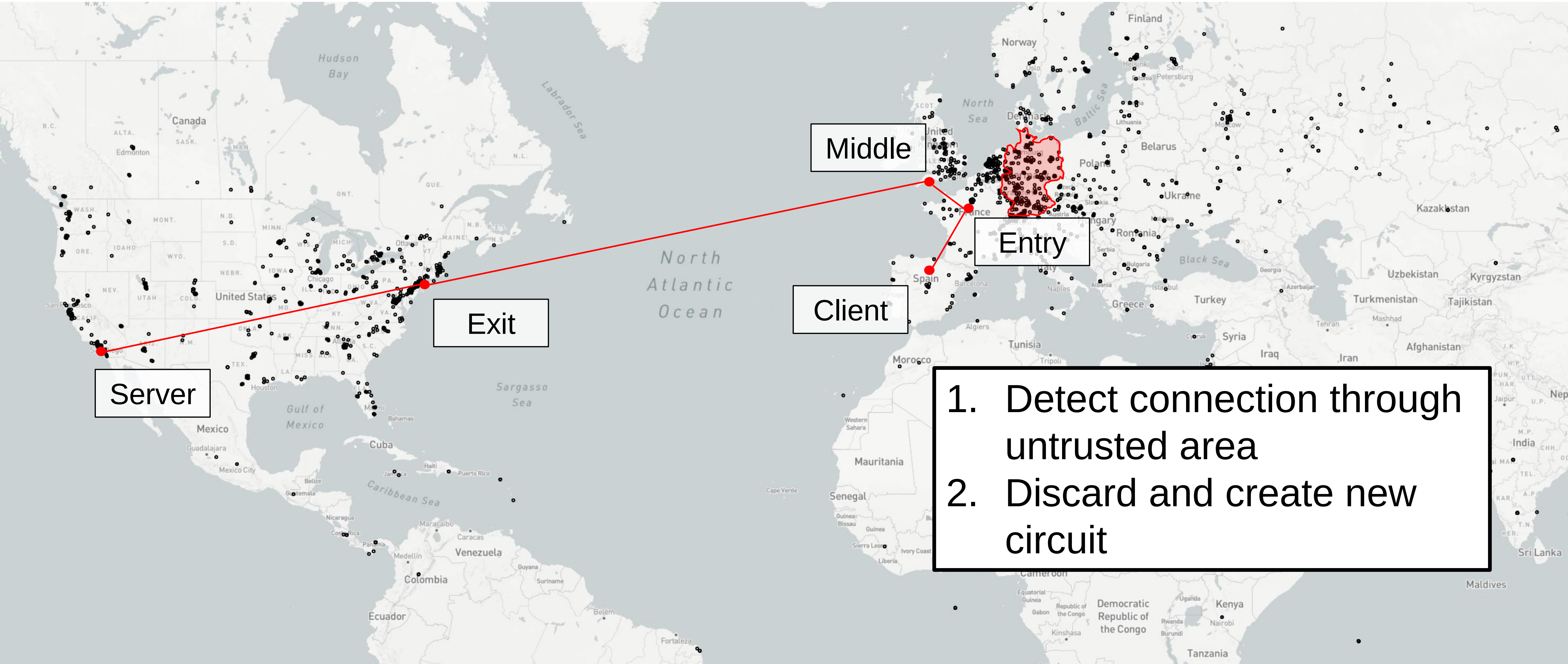
Server

Detect Untrusted Area



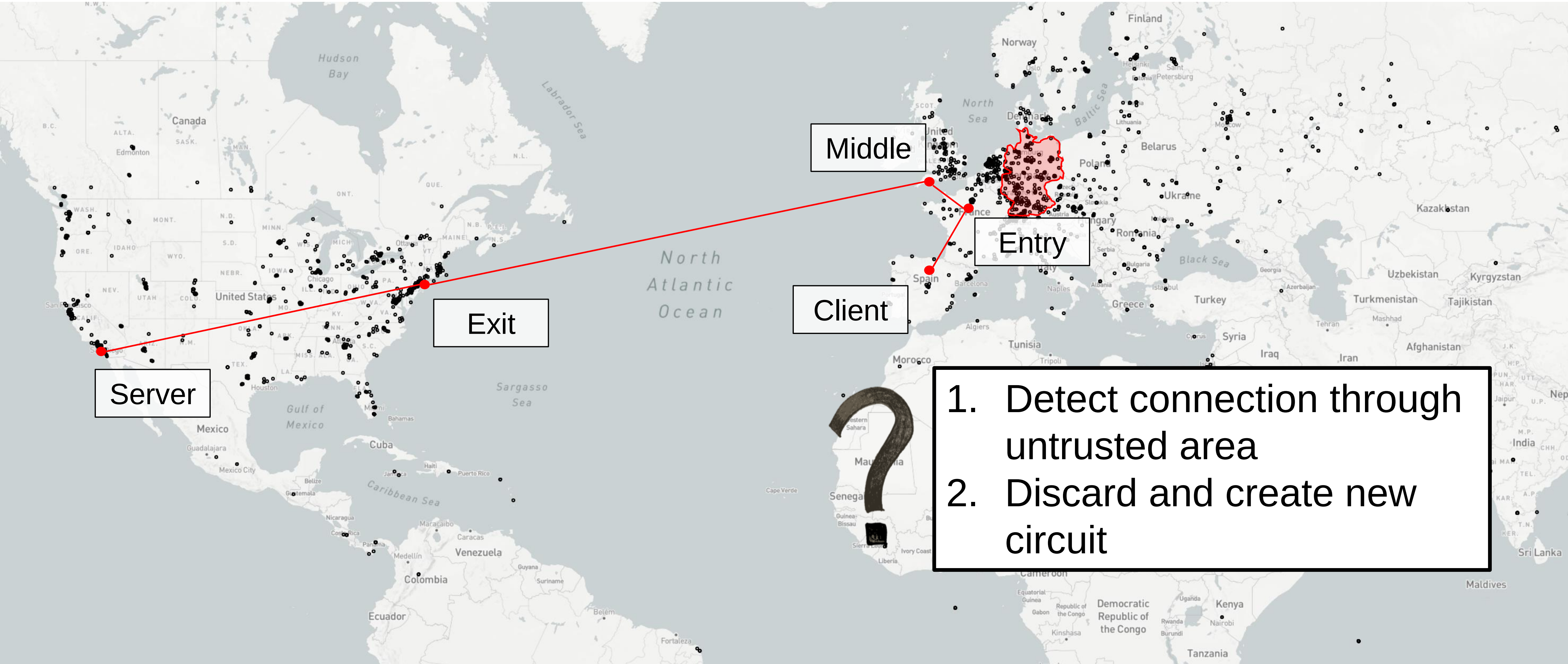
1. Detect connection through untrusted area

Use Better Circuit



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

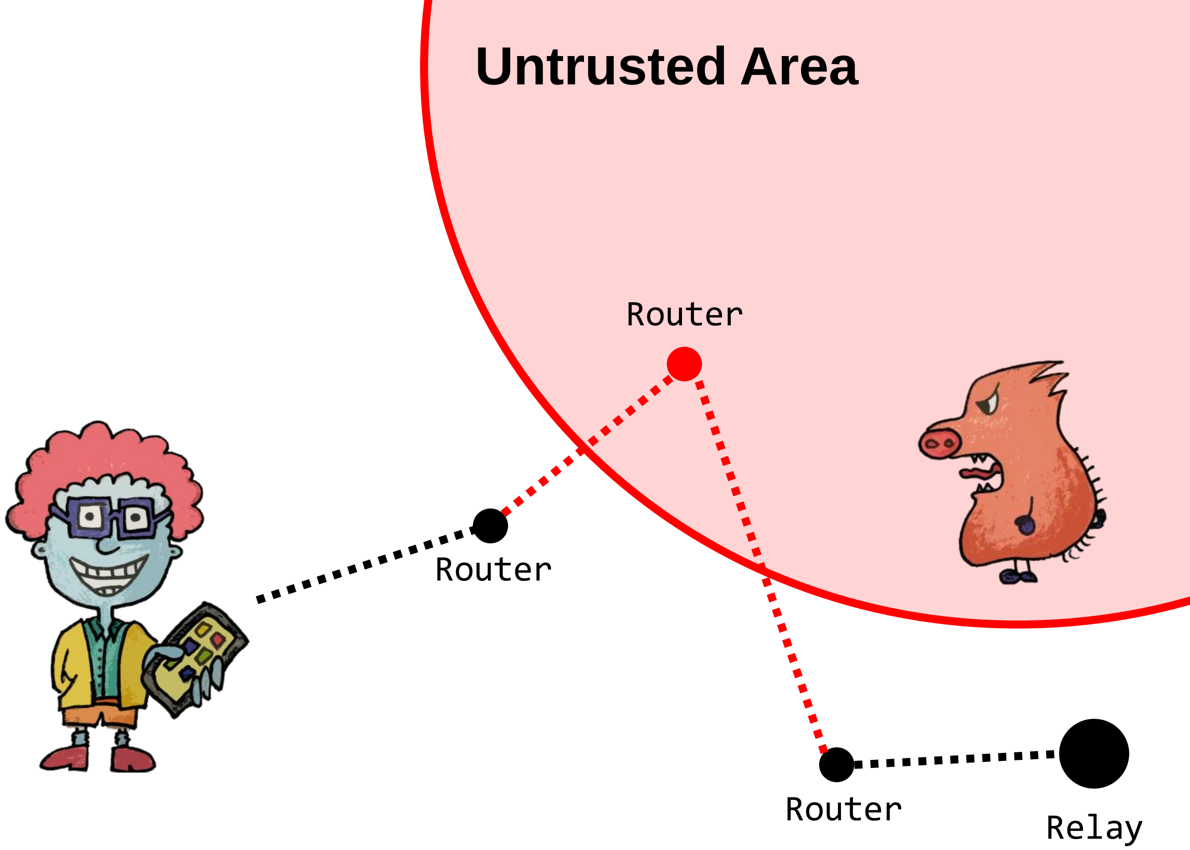
How can we do this?



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

Timing Decisions

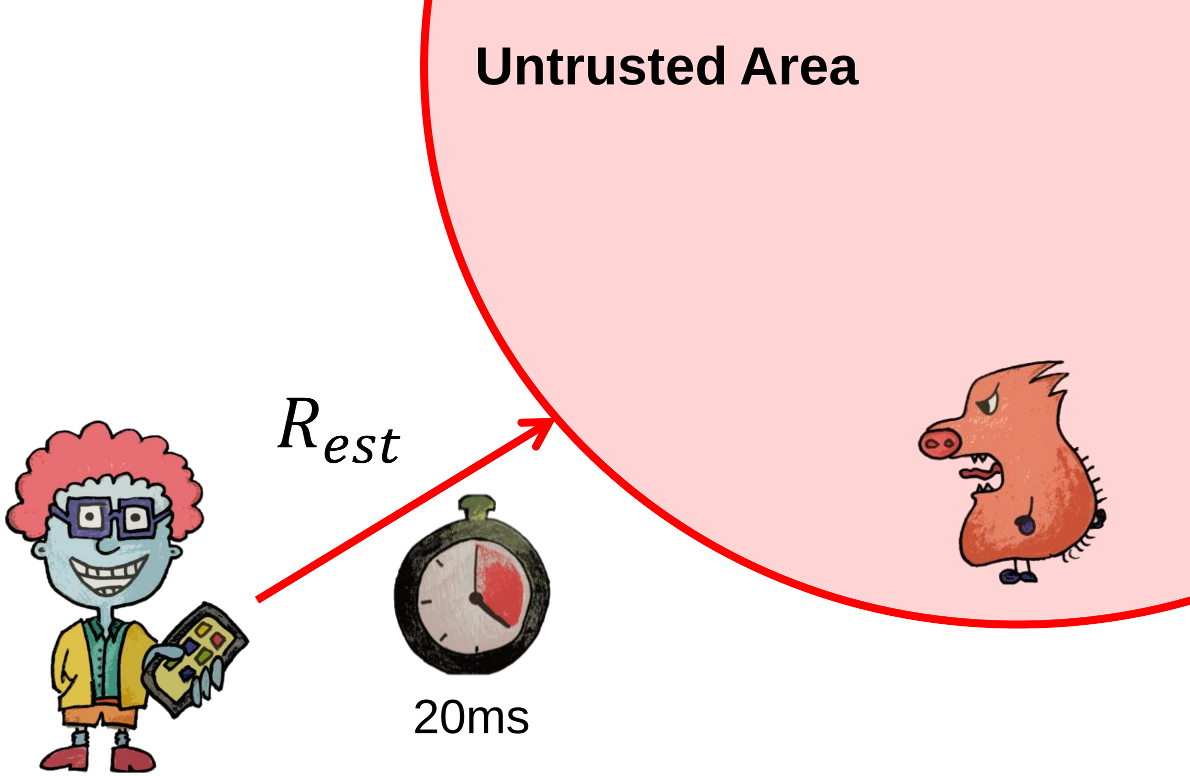
- Detect connection through untrusted area
  - Relays: GeoIP location data
  - Routing: Not transparent
  - → **Measure end-to-end timing**



1. D. Levin, Y. Lee, L. Valenta, Z. Li, V. Lai, C. Lumezanu, N. Spring, and B. Bhattacharjee, "**Alibi Routing**," in *Conference of the ACM Special Interest Group on Data Communication,* SIGCOMM'15
2. Z. Li, S. Herwig, and D. Levin, "**DeTor: Provably Avoiding Geographic Regions in Tor**," in *USENIX Security Symposium,* USENIX'17

Estimate Worst Case

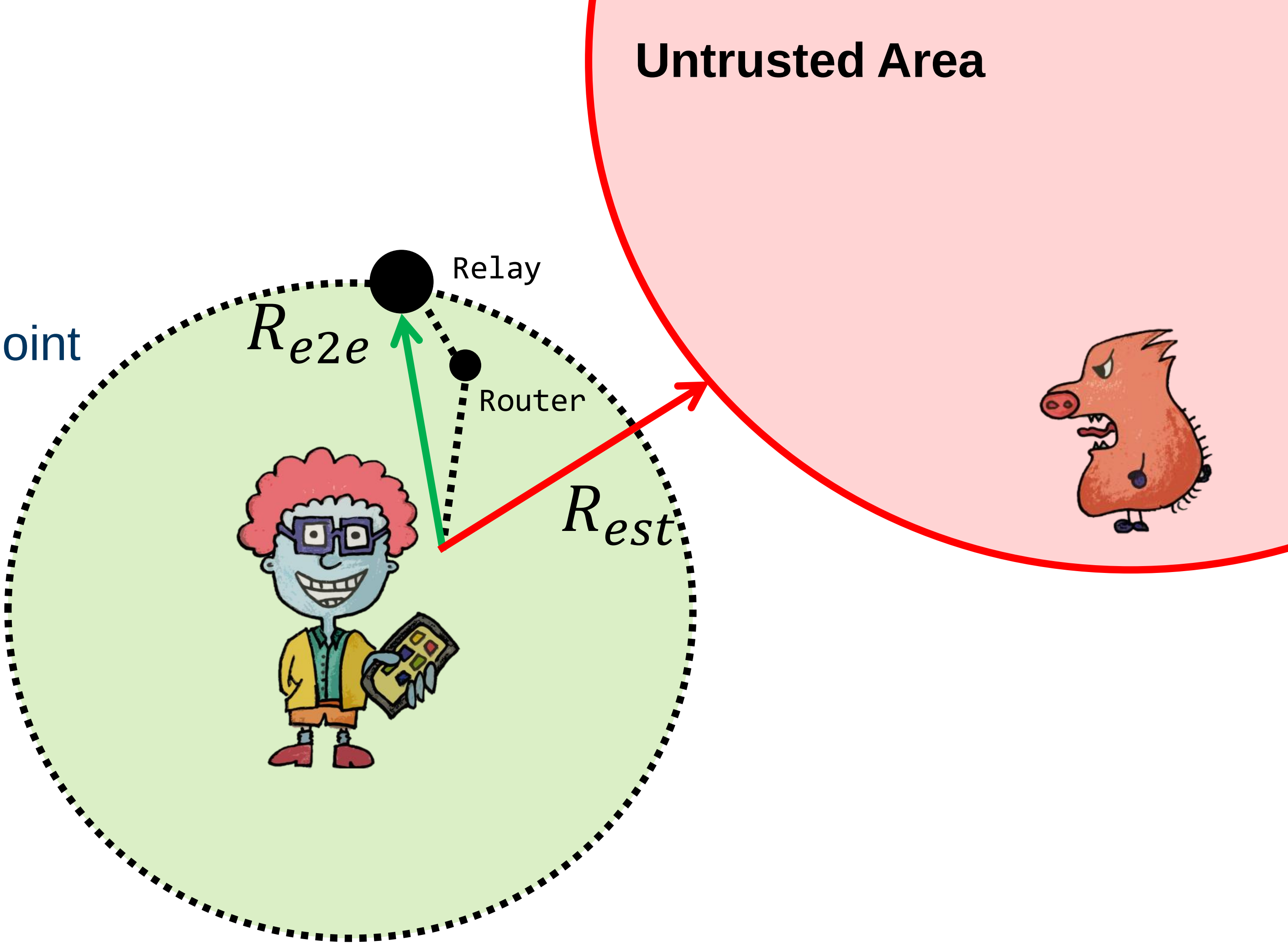
1. Find closest point in untrusted area
2. Measure distance between client and point
3. Assume speed, e.g., $^2/_3$ speed of light
4. Estimate RTT



Timing Decision

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

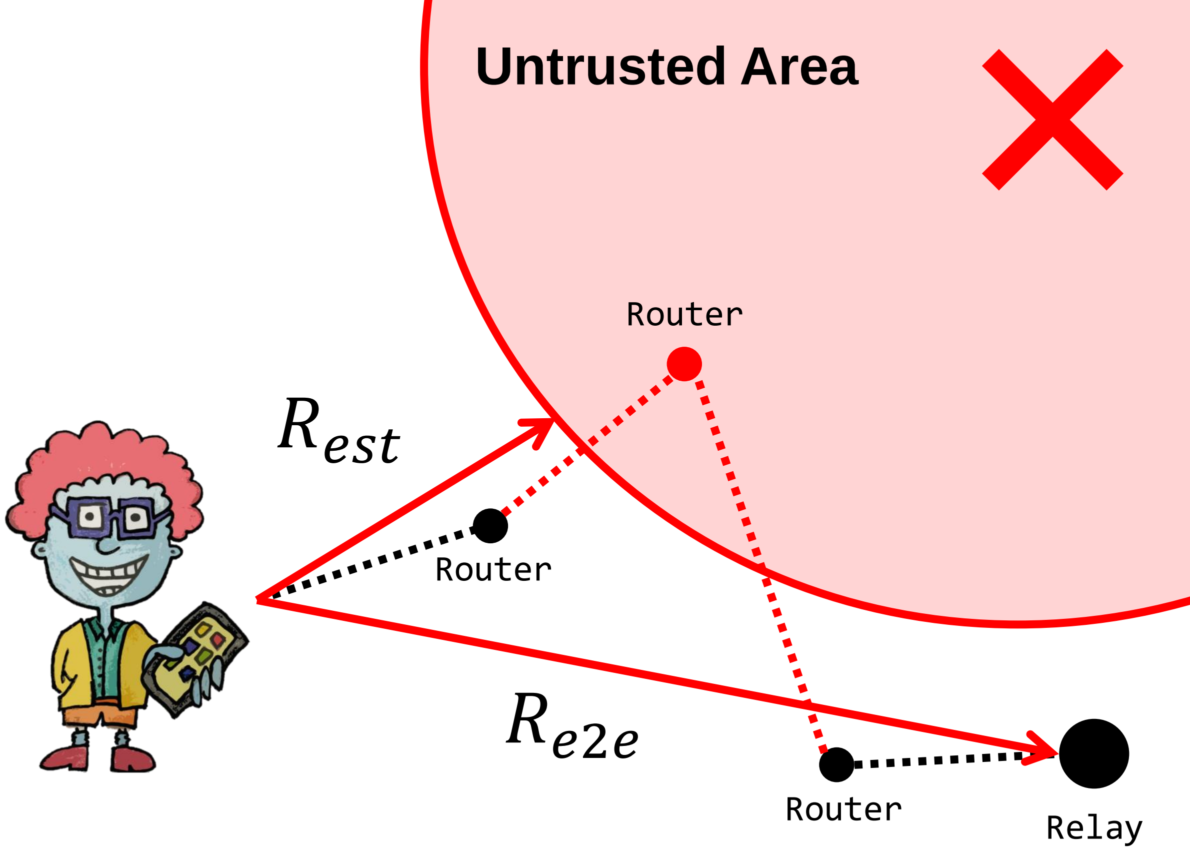
- Use threshold for decisions
  - $R_{e2e} < R_{est}$ ✓



Timing Decision

1. Find closest point in untrusted area
2. Measure distance between client and point
3. Assume speed, e.g., $^2/_3$ speed of light
4. Estimate RTT

- Use threshold for decisions
  - $R_{e2e} < R_{est}$ ✓
  - $R_{e2e} \geq R_{est}$ ✗

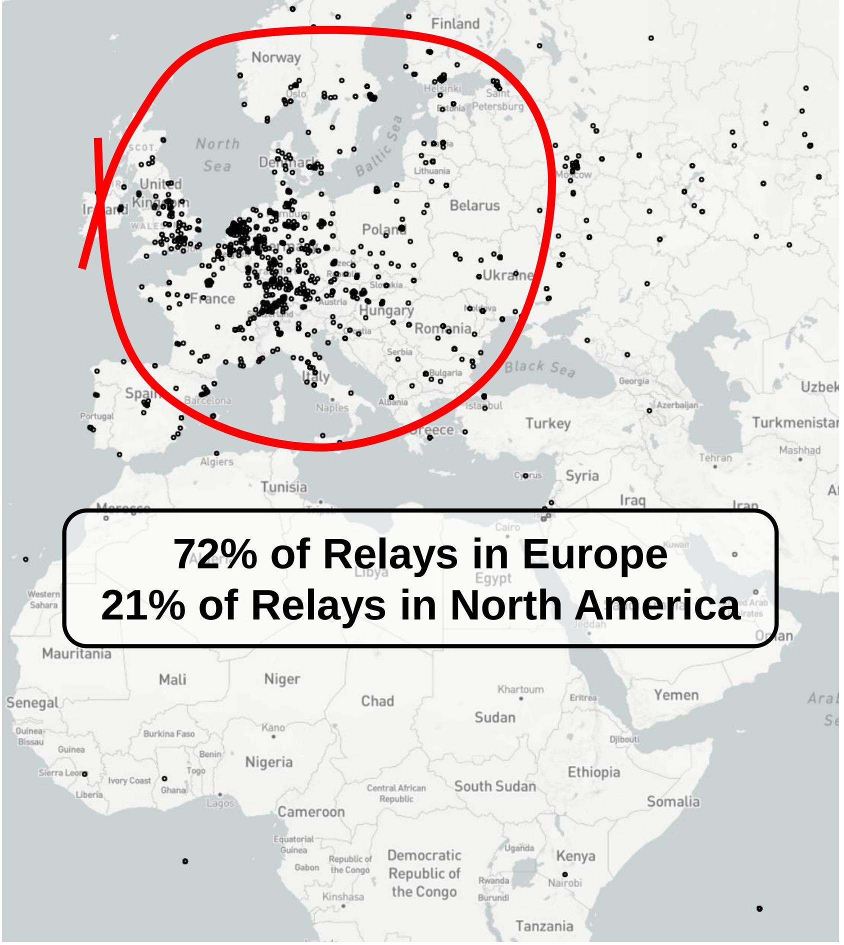


Challenges of Geo Avoidance

Considerations for the system design.

Three Classes of Challenges

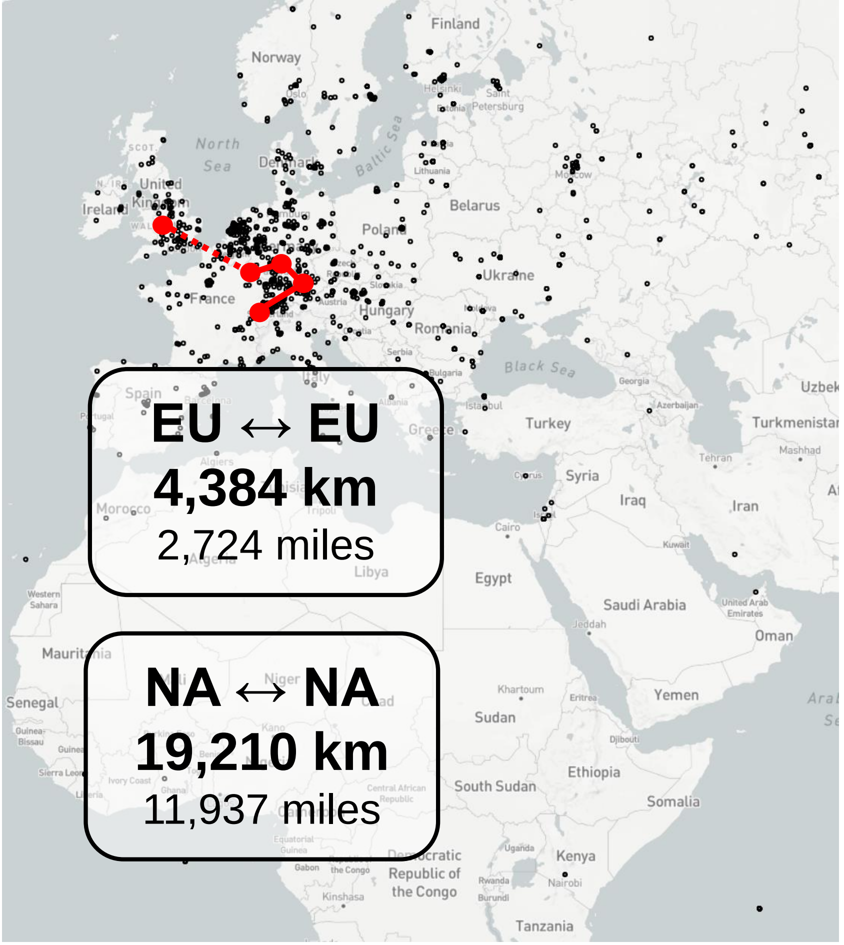
1. **Network Diversity**
   1. *Distribution of Relays*
   2. Varying Connections Lengths
   3. Connection Failures
2. **Ground Truth**
   1. GeoIP Location Errors
   2. Assymetric Routes
   3. Intransparent Transmission Characteristics
3. **Deployment**
   1. Maintaining Tor‘s Performance and Security
   2. Using Reliable Information Sources



72% of Relays in Europe
21% of Relays in North America

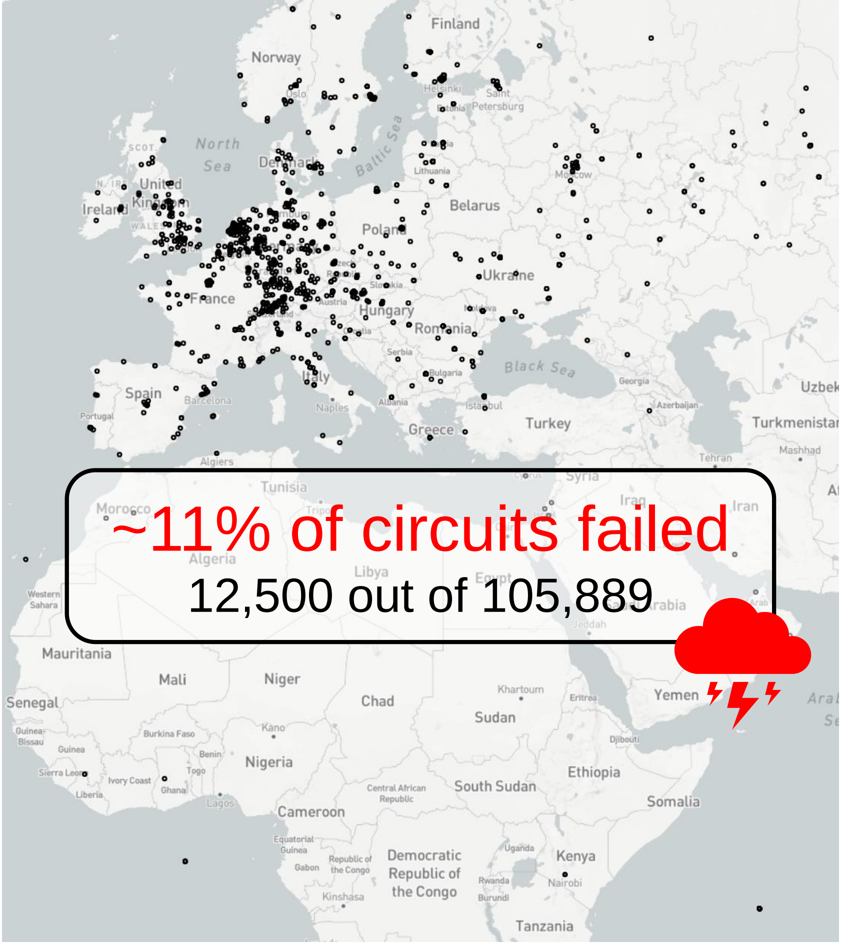
Three Classes of Challenges

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



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

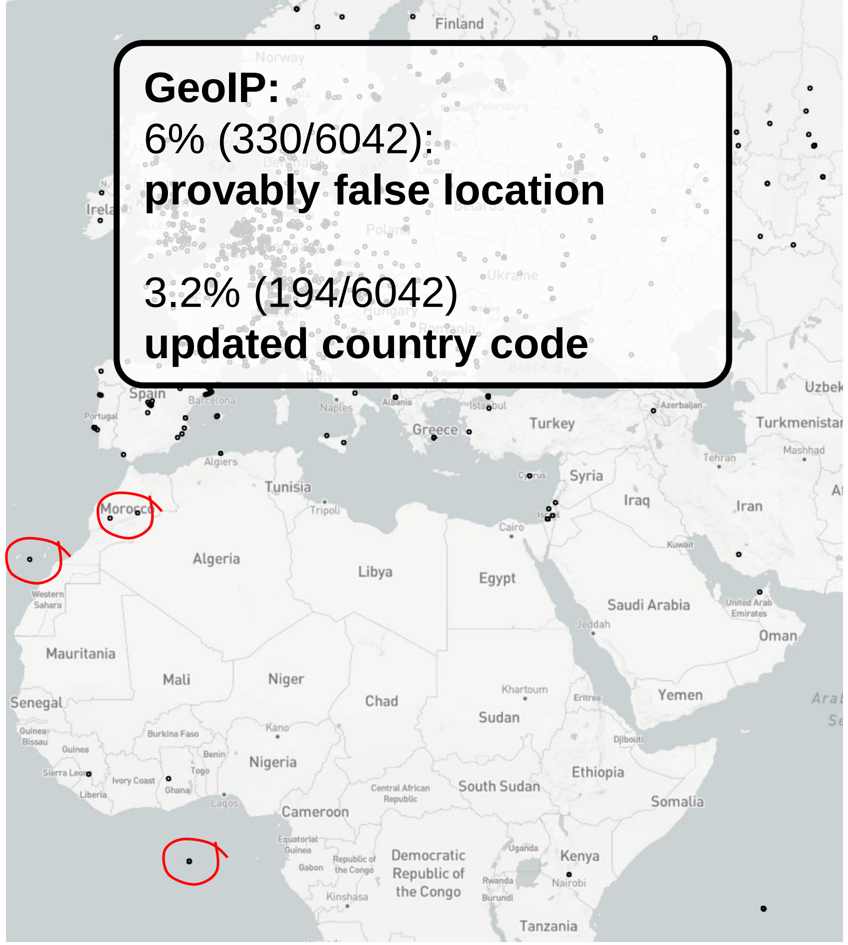


~11% of circuits failed

12,500 out of 105,889

Three Classes of Challenges

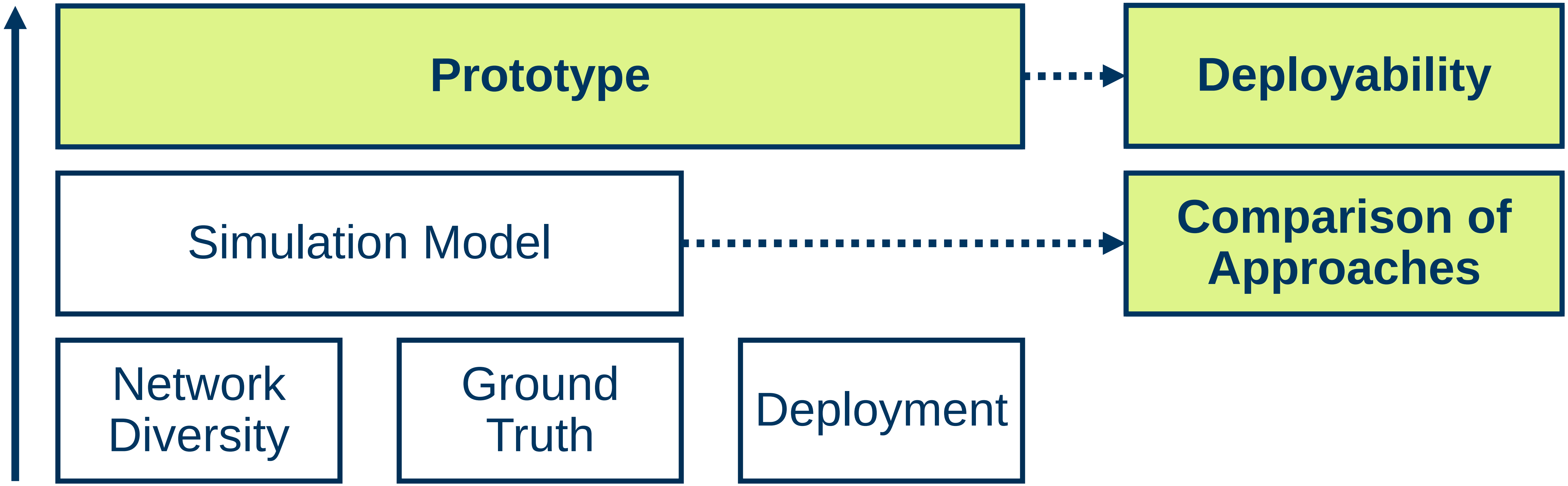
1. **Network Diversity**
   1. Distribution of Relays
   2. Varying Connections Lengths
   3. Connection Failures
2. **Ground Truth**
   1. *GeoIP Location Errors*
   2. Assymetric Routes
   3. Intransparent Transmission Characteristics
3. **Deployment**
   1. Maintaining Tor‘s Performance and Security
   2. Using Reliable Information Sources



GeoIP:
6% (330/6042):
provably false location

3.2% (194/6042)
updated country code

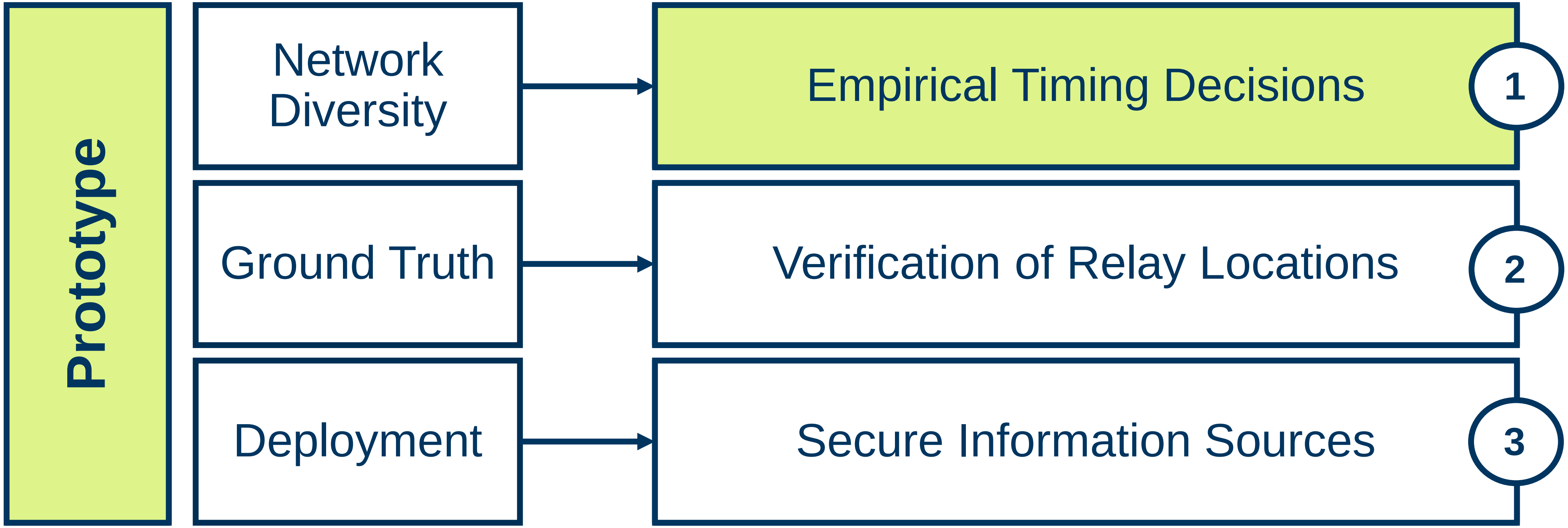
Designing the Avoidance System



Prototype: TrilateraTor

Considering the challenges.

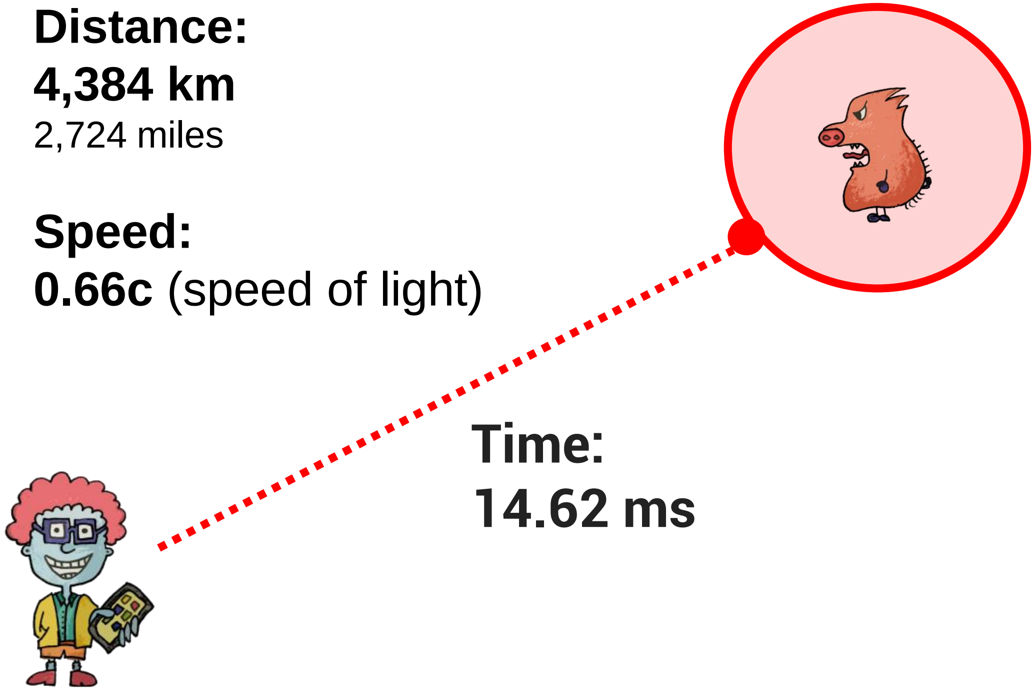
Considering the Challenges


Prototype
Network Diversity
Ground Truth
Deployment
Empirical Timing Decisions
1
Verification of Relay Locations
2
Secure Information Sources
3

Network Diversity: Timing Decisions

1

Upper Bound Decision



Distance:
4,384 km
2,724 miles

Speed:
0.66c (speed of light)

Time:
14.62 ms

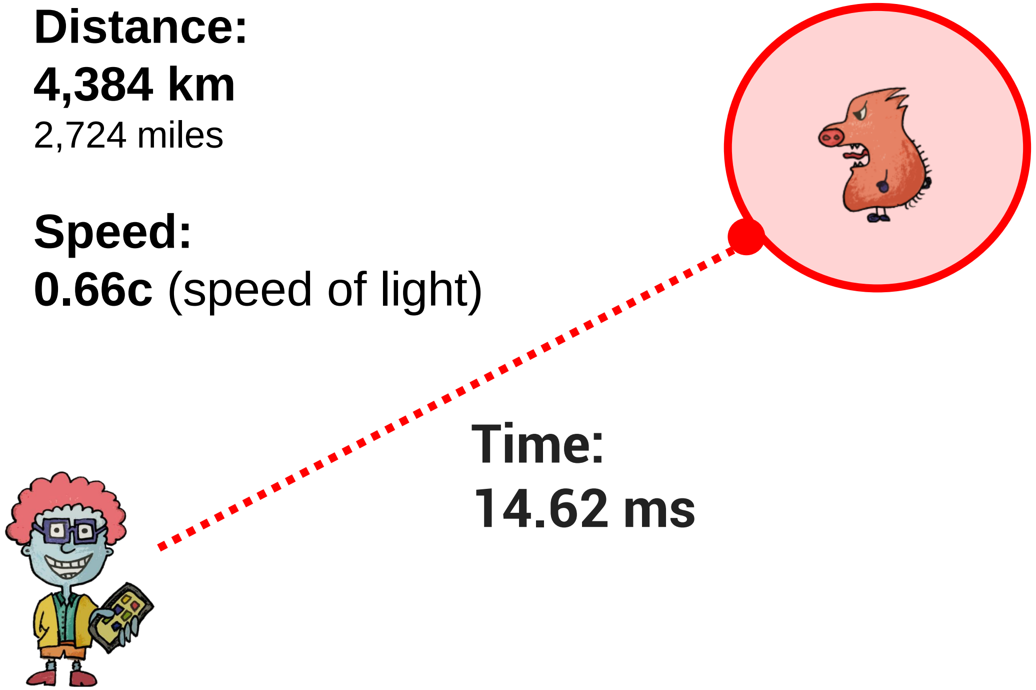
Empirical Timing Decisions

1

Upper Bound Decision

Distance:
4,384 km
2,724 miles

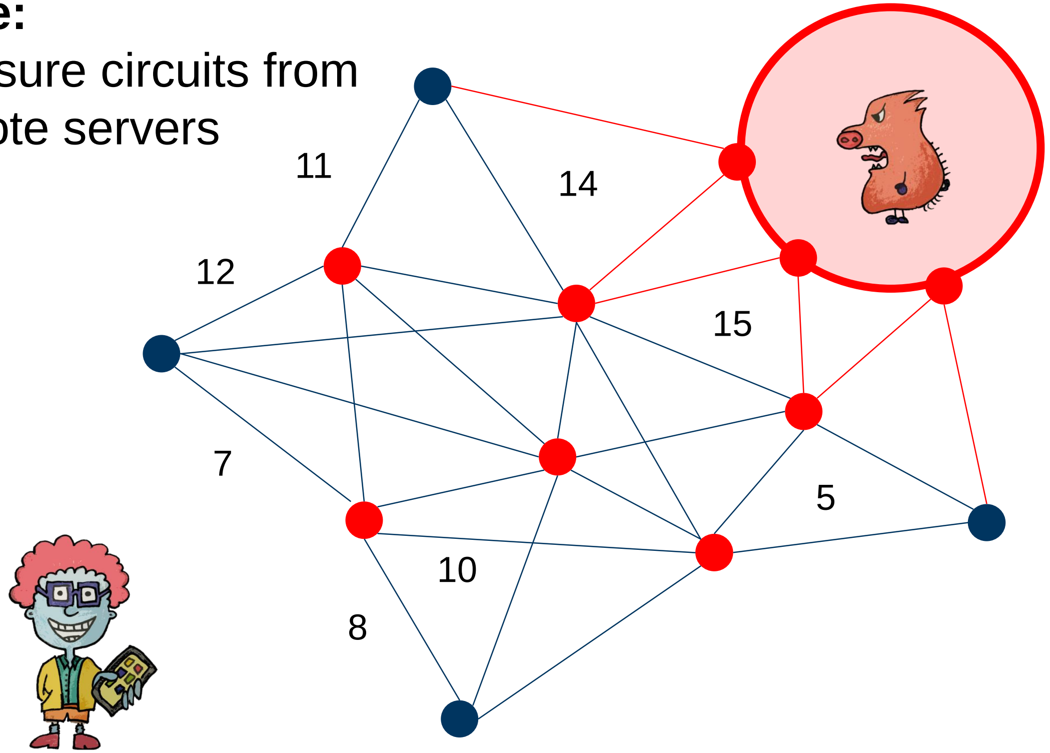
Speed:
0.66c (speed of light)



Time:
14.62 ms

TrilateraTor

Time:
Measure circuits from remote servers



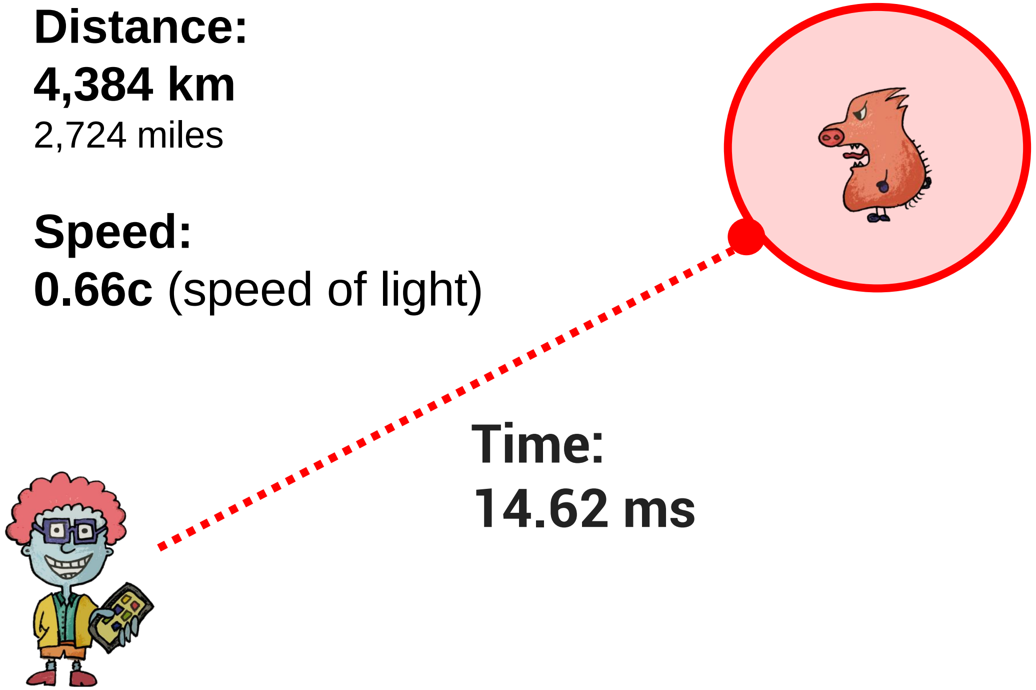
Hop Relations Table

1

Upper Bound Decision

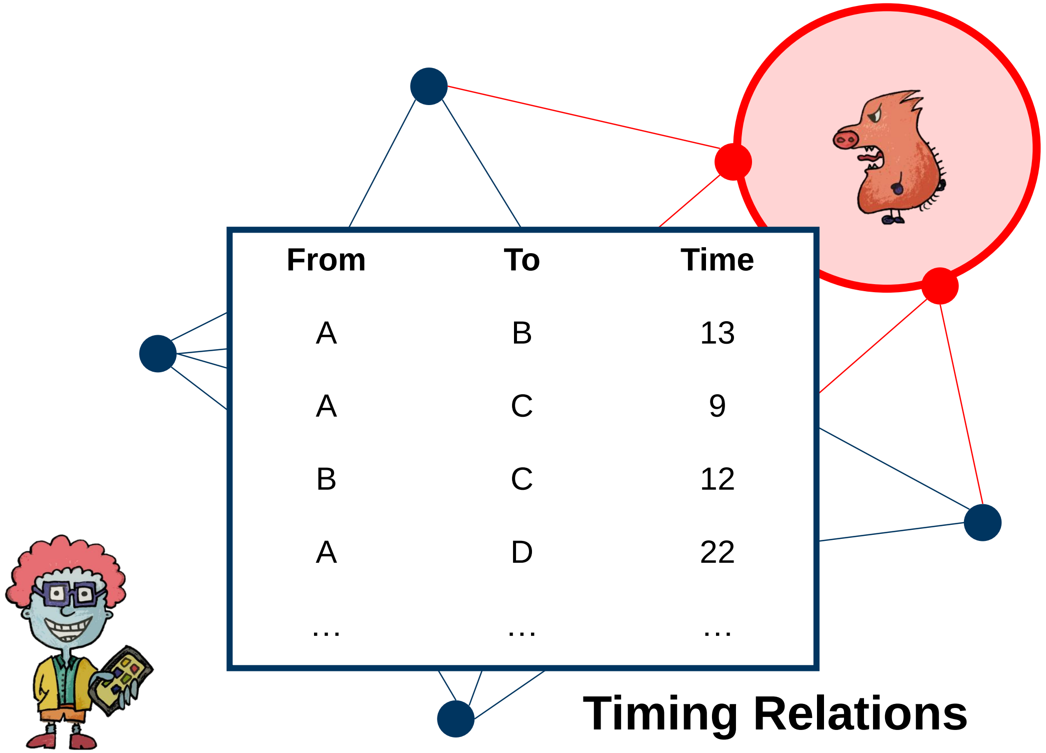
Distance:
4,384 km
2,724 miles

Speed:
0.66c (speed of light)



Time:
14.62 ms

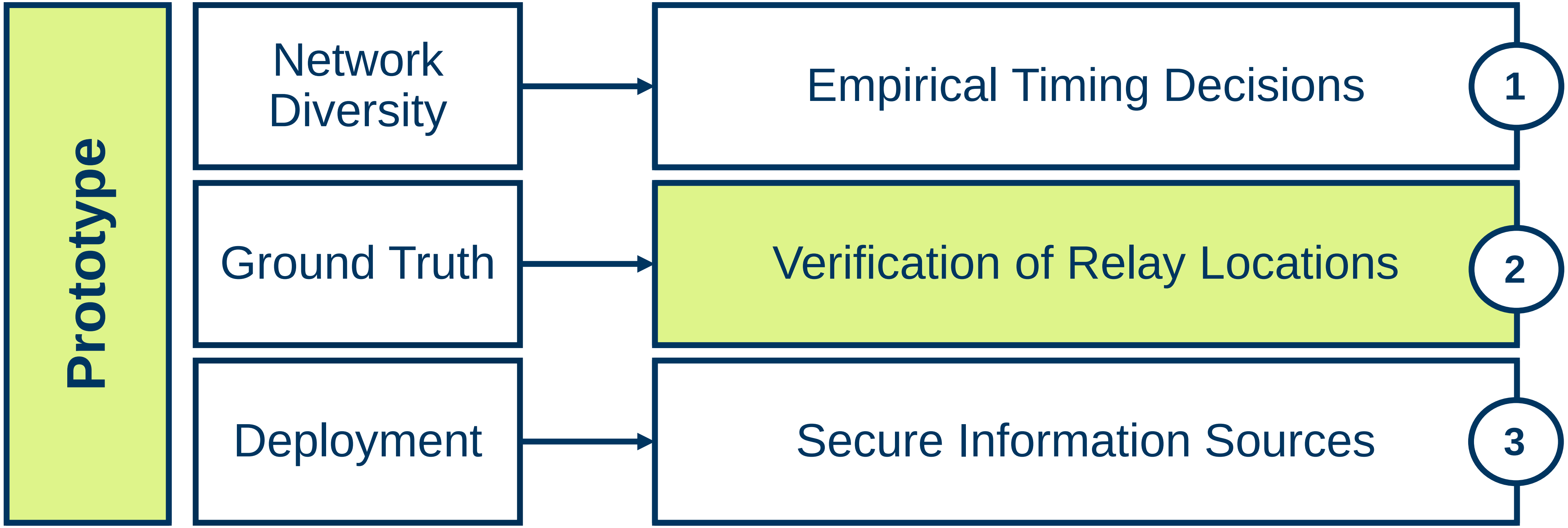
TrilateraTor



From	To	Time
A	B	13
A	C	9
B	C	12
A	D	22
…	…	…

Timing Relations

Considering the Challenges



Prototype

Network Diversity → Empirical Timing Decisions 1

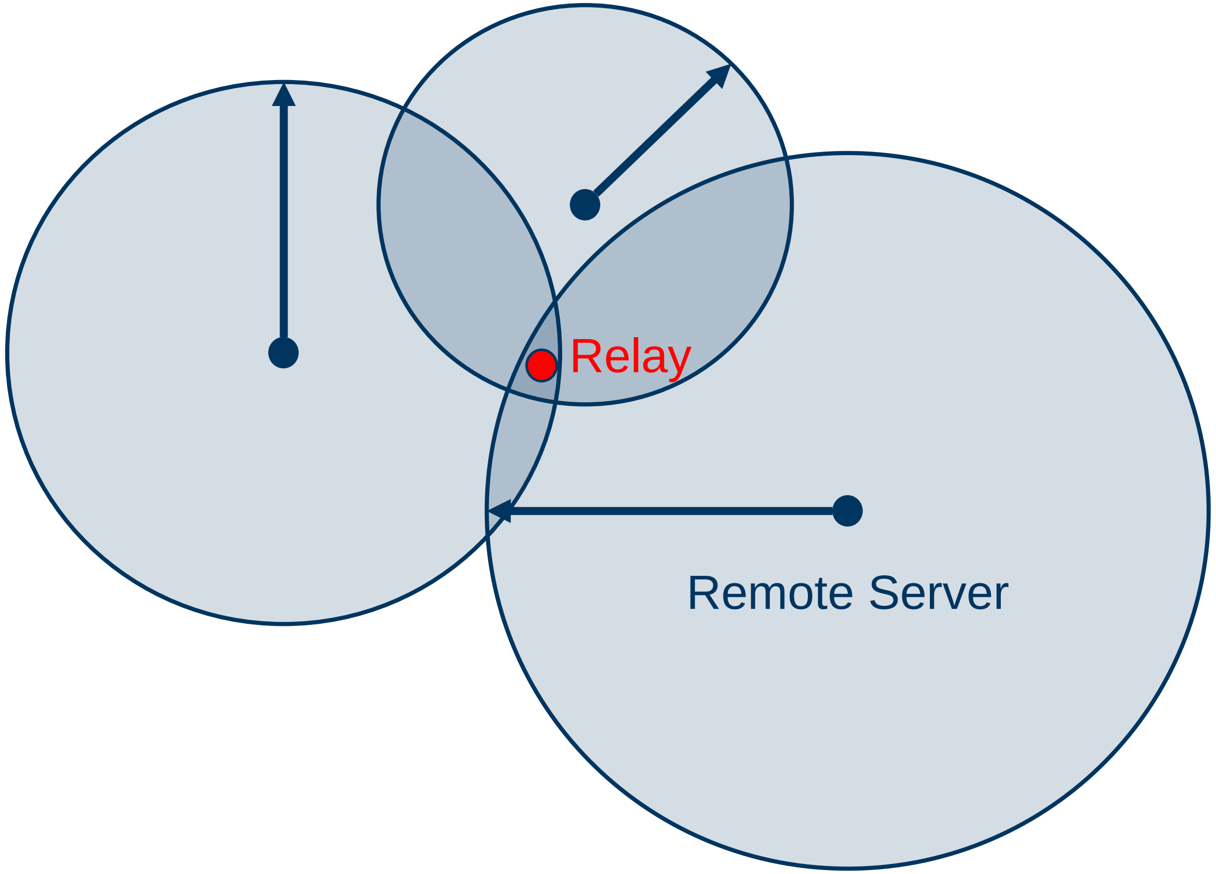
Ground Truth → Verification of Relay Locations 2

Deployment → Secure Information Sources 3

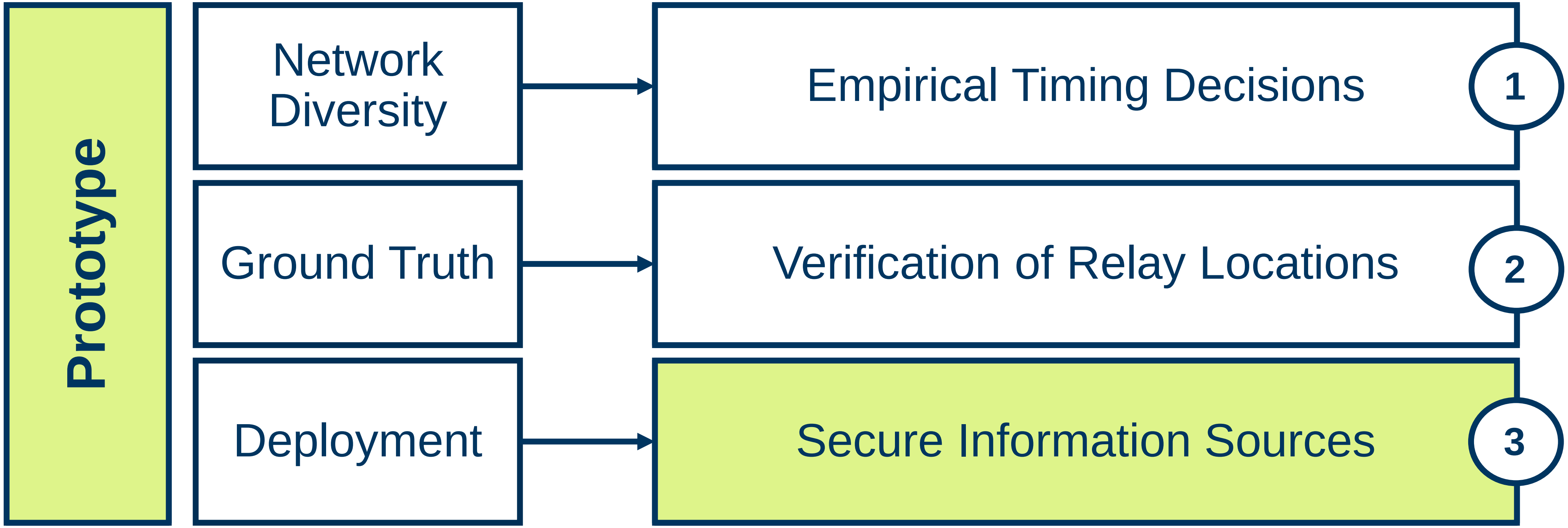
Ground Truth: Relay Locations

2

- 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

1

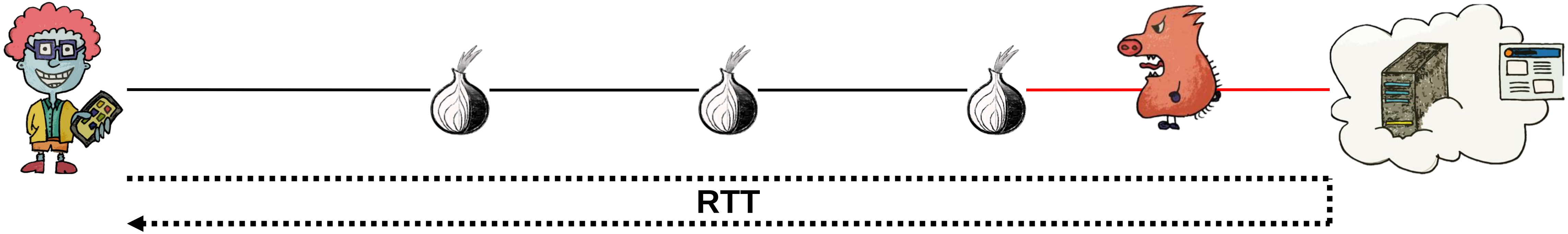
Verification of Relay Locations

2

Secure Information Sources

3

Deployment: Timing Measurements

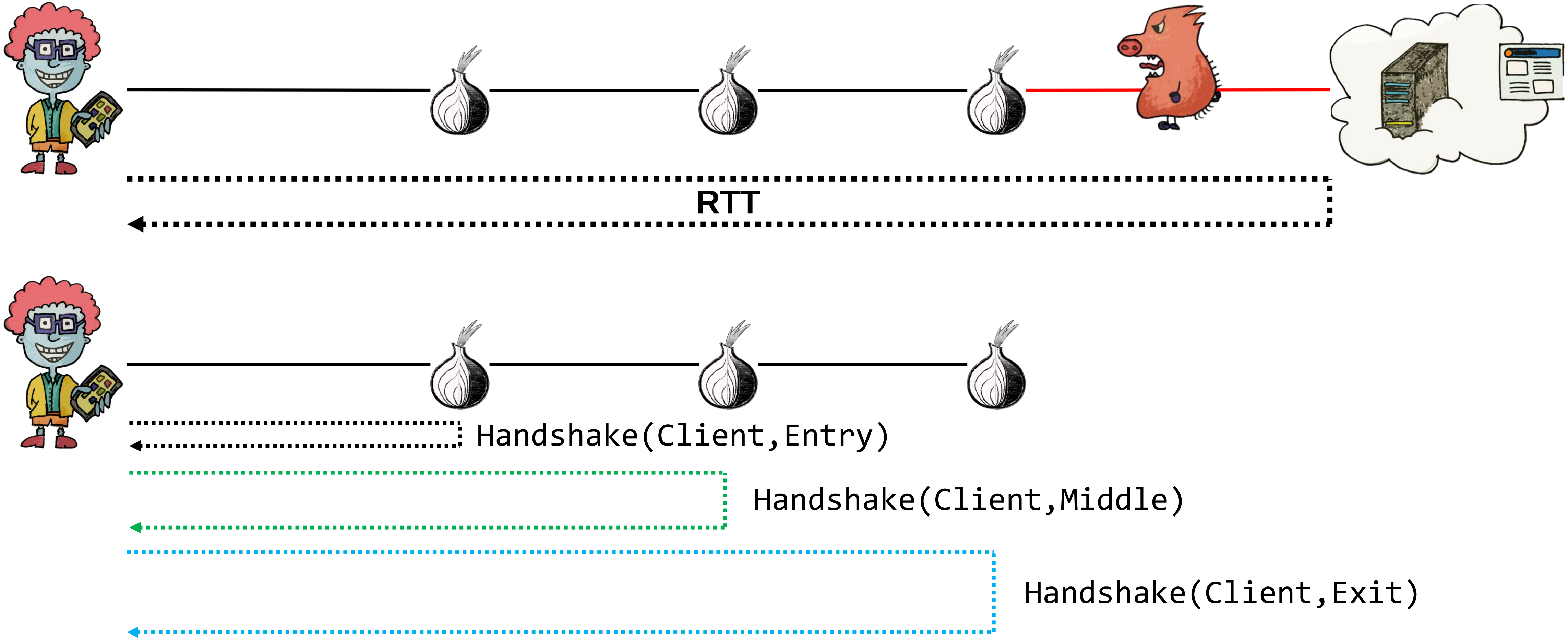


- Prior work: Probe the entire circuit
- Circuit is not checked at this point
- Two major issues:
  - Security: Reveals endpoint to adversary
  - Performance: Requires additional measurements

1. D. Levin, Y. Lee, L. Valenta, Z. Li, V. Lai, C. Lumezanu, N. Spring, and B. Bhattacharjee, "**Alibi Routing**," in *Conference of the ACM Special Interest Group on Data Communication,* SIGCOMM'15
2. Z. Li, S. Herwig, and D. Levin, "**DeTor: Provably Avoiding Geographic Regions in Tor**," in *USENIX Security Symposium,* USENIX'17

Alternative: Handshake Timings

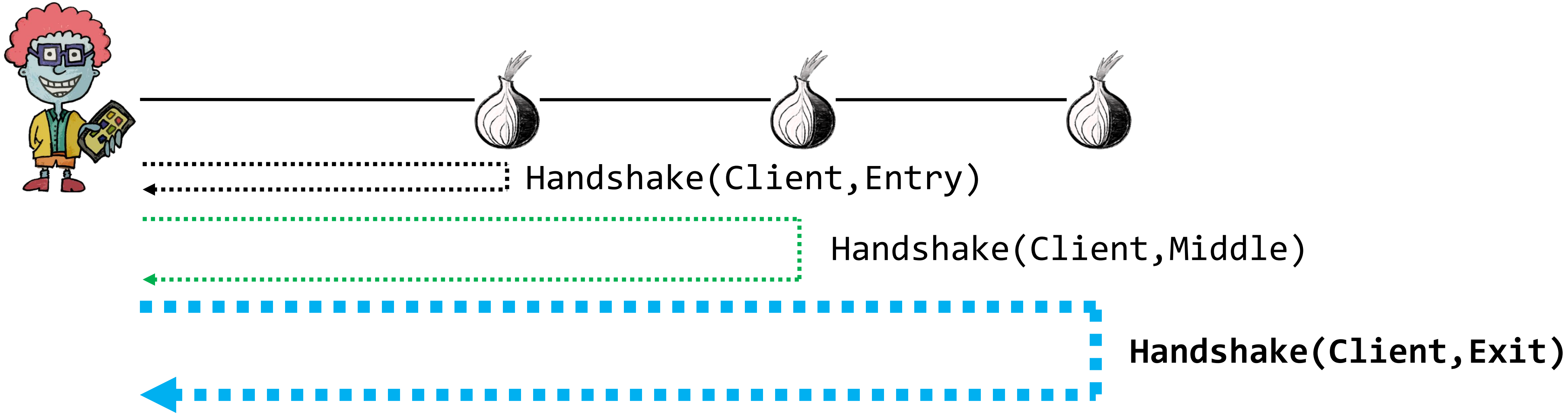
3



Secure Information Sources

3

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

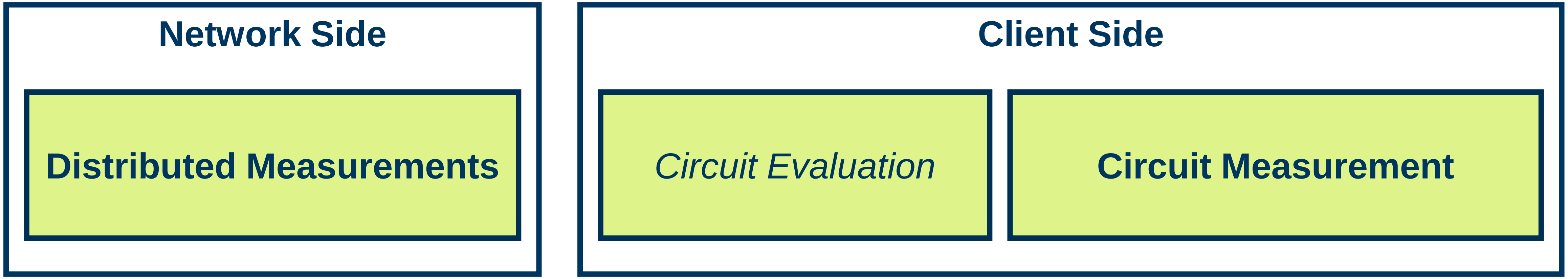


Technical Concept

Network Side

Client Side

Two Types of Measurements



Network Side

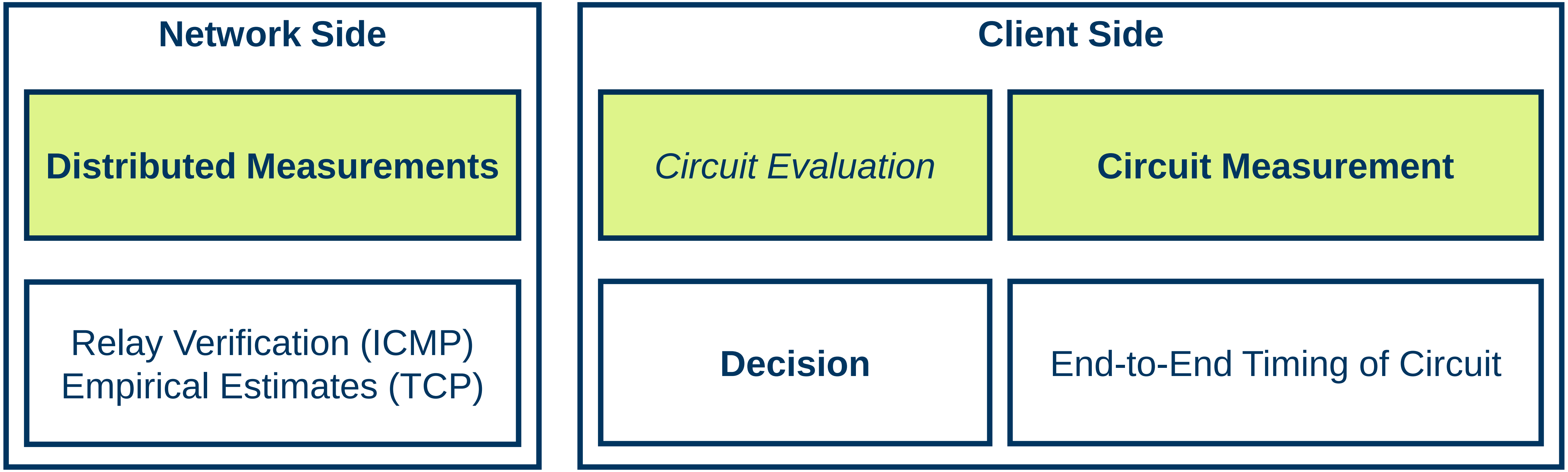
Distributed Measurements

Client Side

Circuit Evaluation

Circuit Measurement

Decision Data



Network Side

Distributed Measurements

Relay Verification (ICMP)
Empirical Estimates (TCP)

Client Side

Circuit Evaluation

Circuit Measurement

Decision

End-to-End Timing of Circuit

$$\boldsymbol{R_{est}} \dashrightarrow \boldsymbol{R_{e2e}} < \boldsymbol{R_{est}}? \dashleftarrow \boldsymbol{R_{e2e}}$$

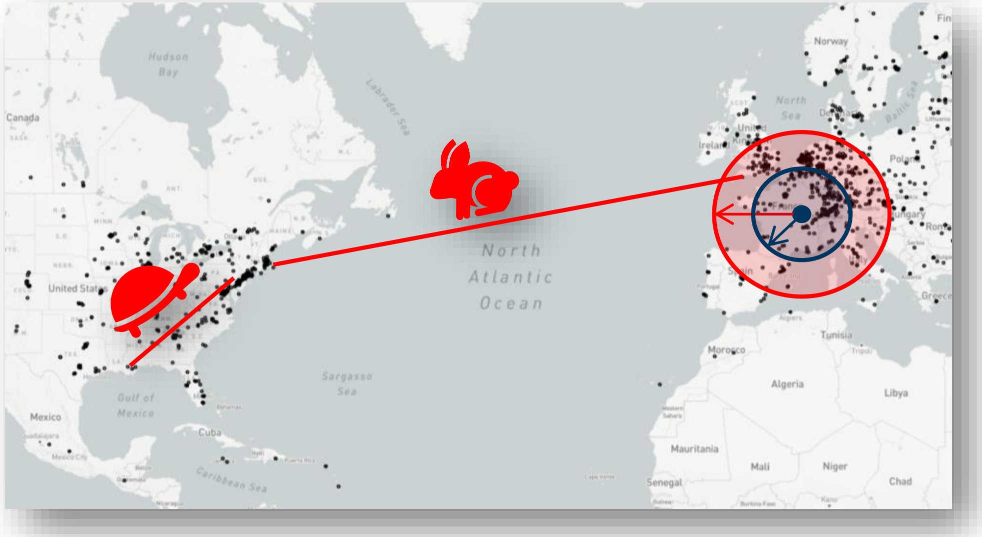
Experiments

Gathering empirical data, comparing approaches.

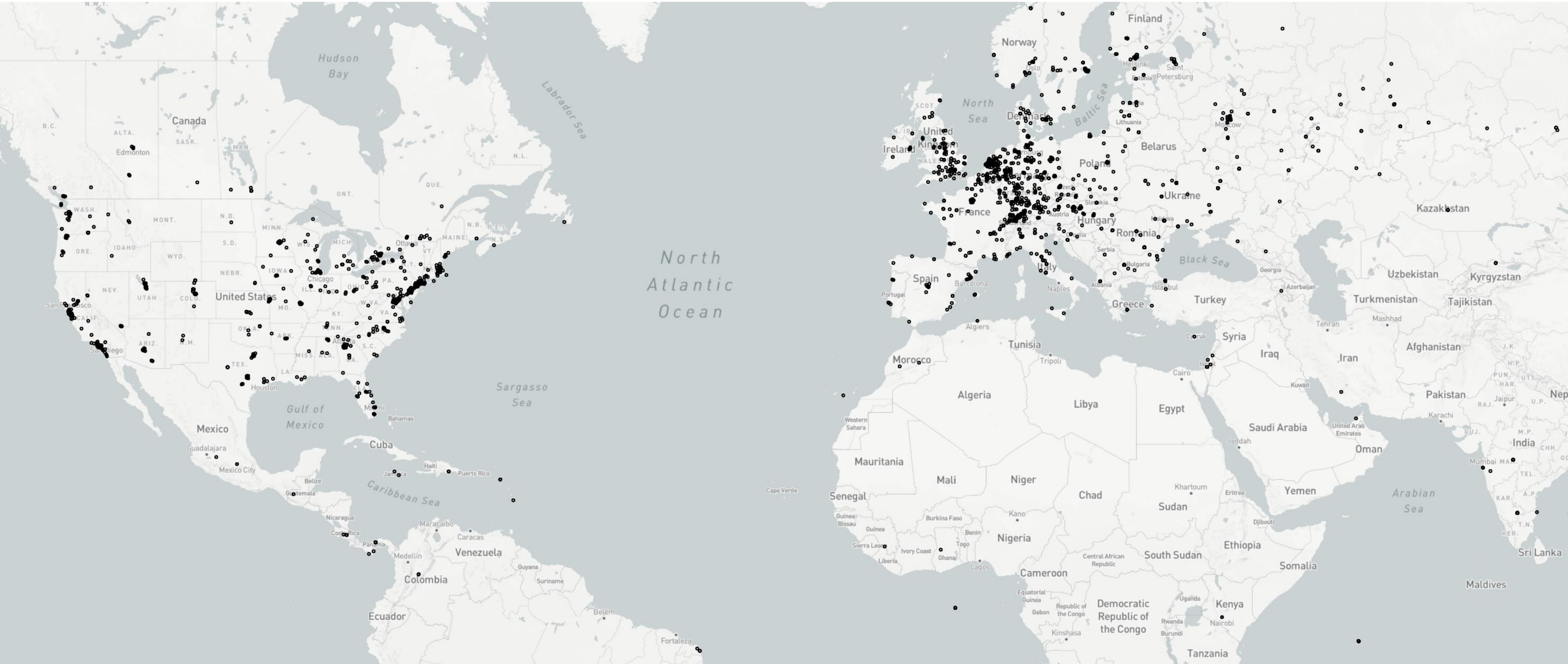
Metrics: How to measure what we achieved

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

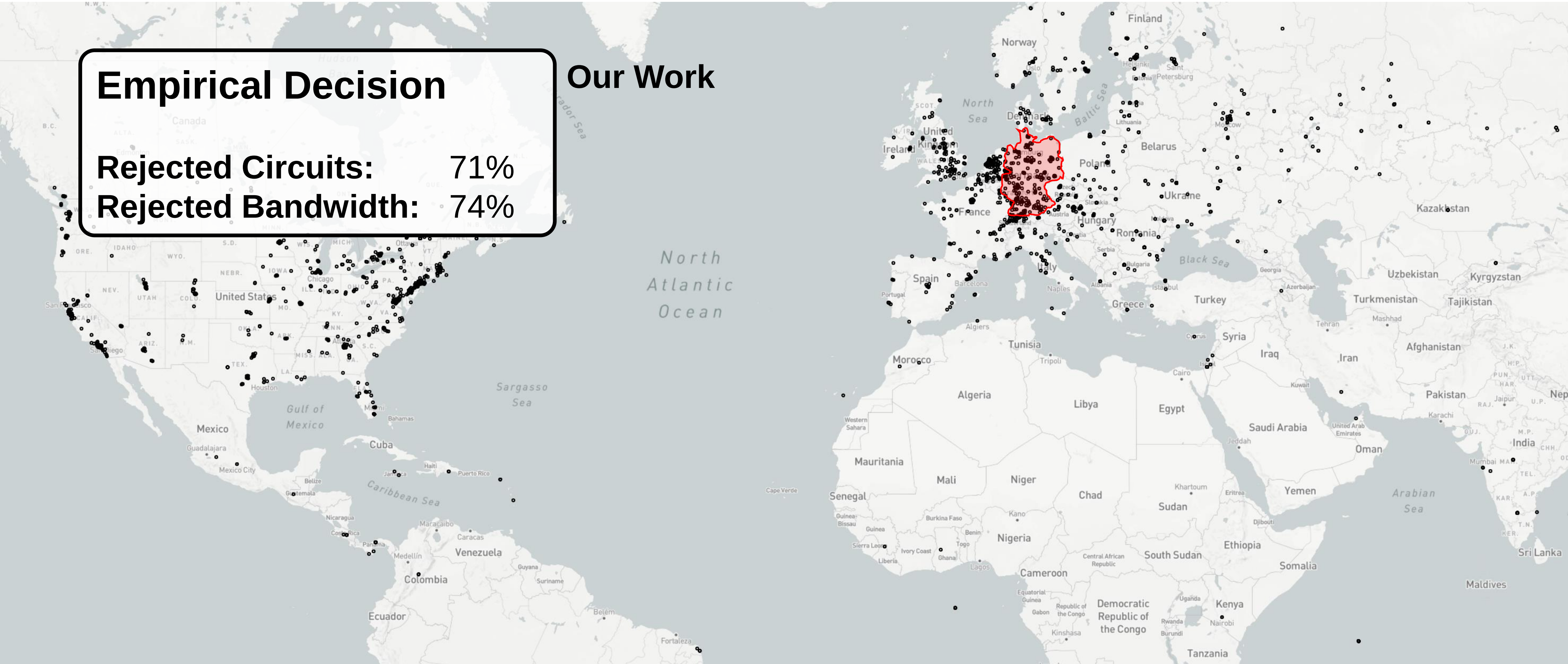
- We measure:
  - **What if…? Loss of bandwidth and circuits in different scenarios.**
  - Time Ratio: Difference between the measured and the estimated time.



What if…?


Hudson Bay
Canada
Labrador Sea
North Atlantic Ocean
United States
Chicago
Ottawa
Houston
Gulf of Mexico
Mexico
Sargasso Sea
Cuba
Caribbean Sea
Venezuela
Colombia
Ecuador
Finland
Norway
North Sea
United Kingdom
Ireland
Denmark
Baltic Sea
Poland
Belarus
Ukraine
France
Hungary
Romania
Spain
Italy
Greece
Turkey
Black Sea
Kazakhstan
Uzbekistan
Kyrgyzstan
Turkmenistan
Tajikistan
Syria
Iraq
Iran
Afghanistan
Pakistan
Morocco
Tunisia
Algeria
Libya
Egypt
Saudi Arabia
Oman
India
Mauritania
Mali
Niger
Chad
Sudan
Yemen
Arabian Sea
Senegal
Nigeria
Ethiopia
South Sudan
Somalia
Cameroon
Democratic Republic of the Congo
Kenya
Tanzania
Sri Lanka
Maldives

What if Germany was forbidden area?



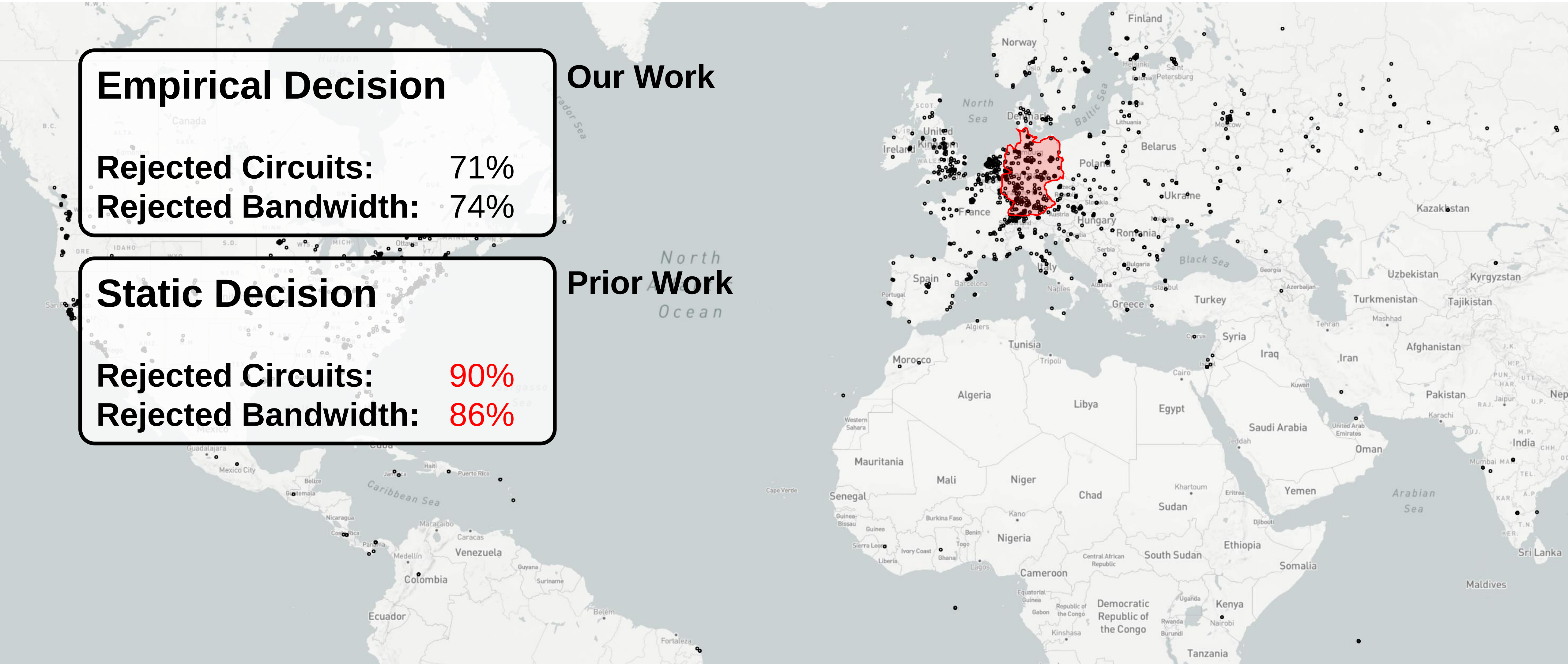
Empirical Decision

Rejected Circuits: 71%

Rejected Bandwidth: 74%

Our Work

What if Germany was forbidden area?



Empirical Decision

Rejected Circuits: 71%

Rejected Bandwidth: 74%

Our Work

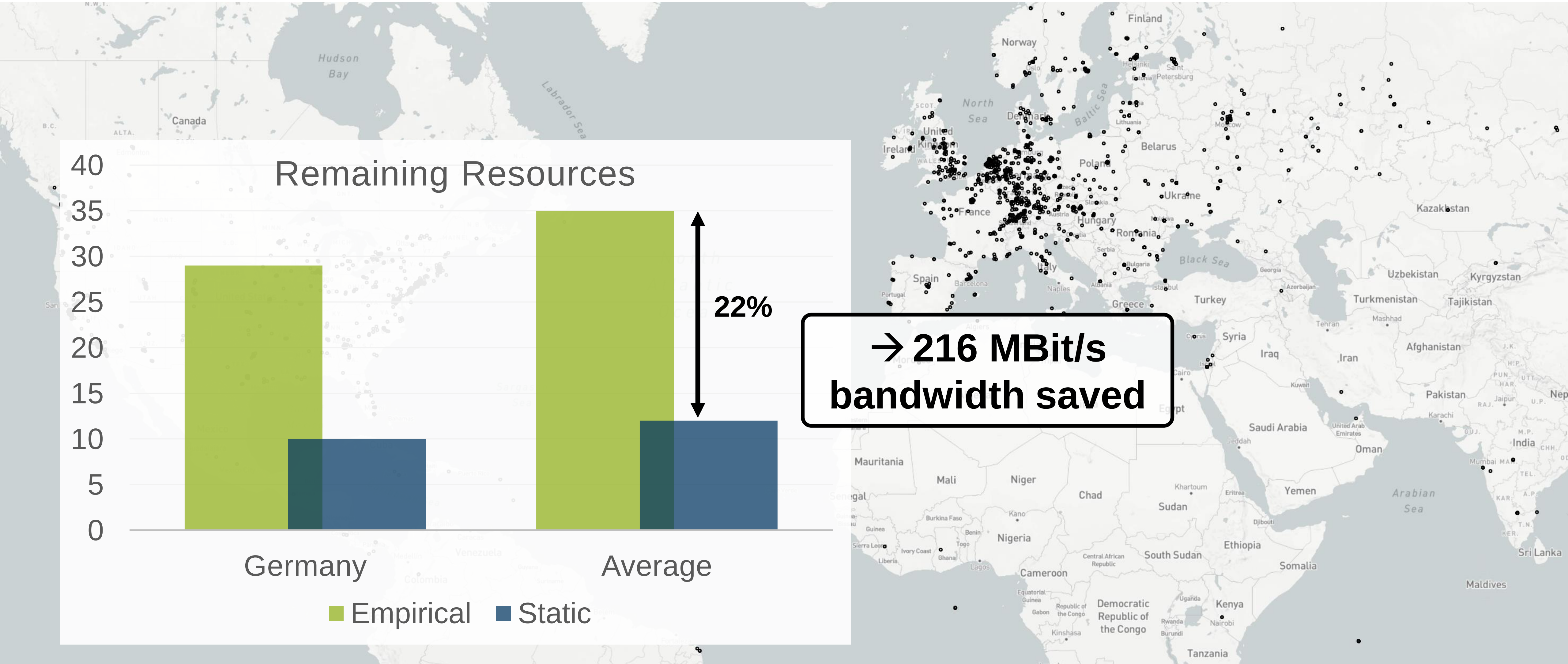
Static Decision

Rejected Circuits: 90%

Rejected Bandwidth: 86%

Prior Work

Limit Performance Impairments



→ 216 MBit/s bandwidth saved

Conclusion

Lessons learned.

Challenges of Geographical Avoidance

3 Classes of Challenges

1. Network Diversity
2. Ground Truth
3. Deployment

Finland
Norway
North Sea
Belarus
Poland
Ukraine
France
Romania
Black Sea
Spain
Greece
Turkey
Turkmenista
Uzbe
Syria
Iraq
Iran
Morocco
Tunisia
Algeria
Libya
Egypt
Saudi Arabia

Designing an Avoidance System

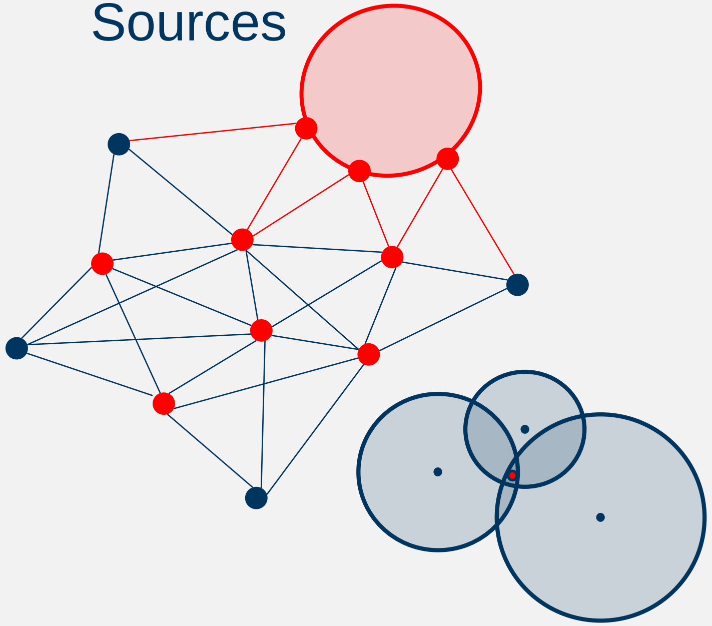
3 Classes of Challenges

1. Network Diversity
2. Ground Truth
3. Deployment

Finland
Norway
North Sea
Ireland
Belarus
Ukraine
France
Spain
Black Sea
Turkey
Greece
Syria
Iraq
Iran
Turkmenista
Uzbe
Tunisia
Morocco
Algeria
Libya
Egypt

Main Features

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



Prototype with Tradeoff

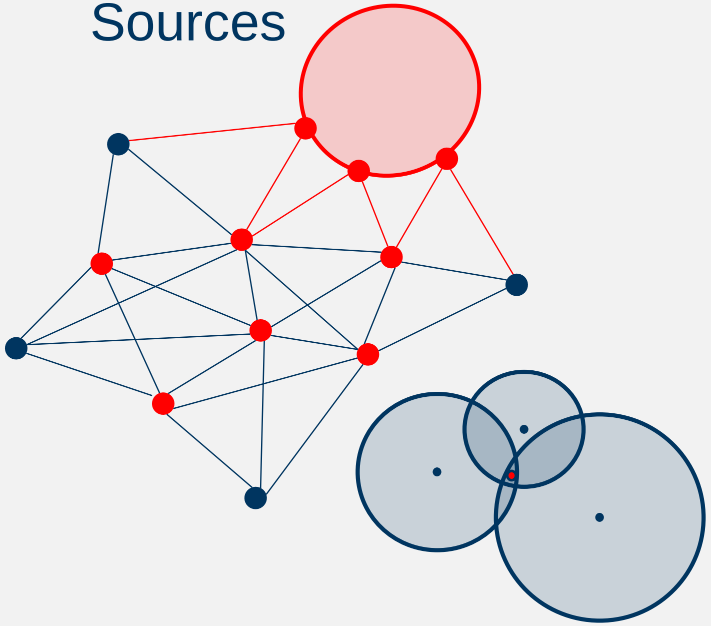
3 Classes of Challenges

1. Network Diversity
2. Ground Truth
3. Deployment

Finland
Norway
North Sea
Denmark
United Kingdom
Ireland
Belarus
Poland
Ukraine
France
Hungary
Romania
Black Sea
Spain
Italy
Greece
Turkey
Uzbe
Turkmenista
Tunisia
Syria
Morocco
Iraq
Iran
Algeria
Libya
Egypt

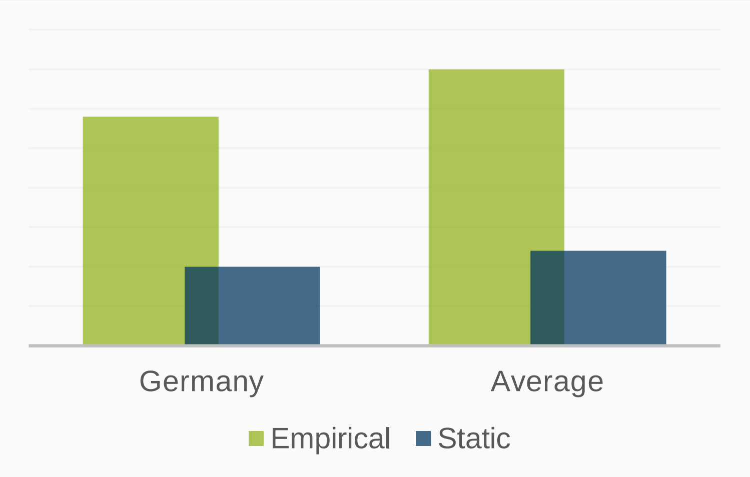
Main Features

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



Evaluation

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



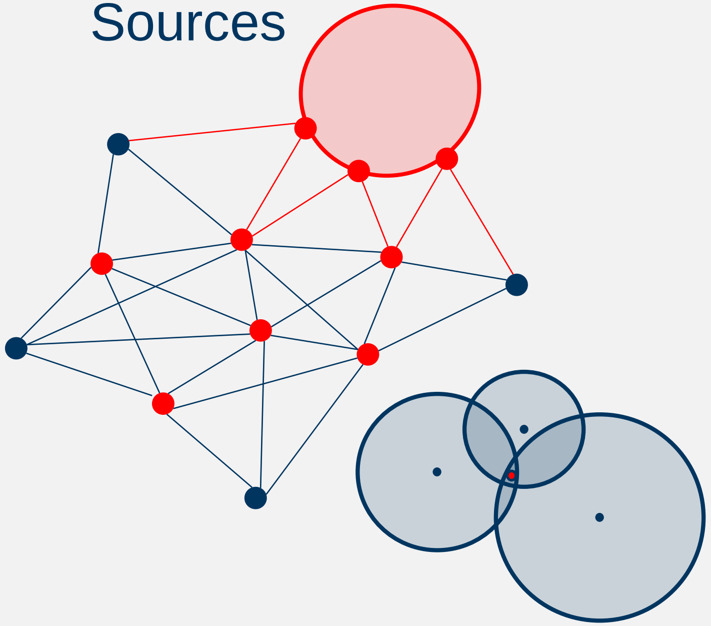
Thank You! Questions?

3 Classes of Challenges

1. Network Diversity
2. Ground Truth
3. Deployment

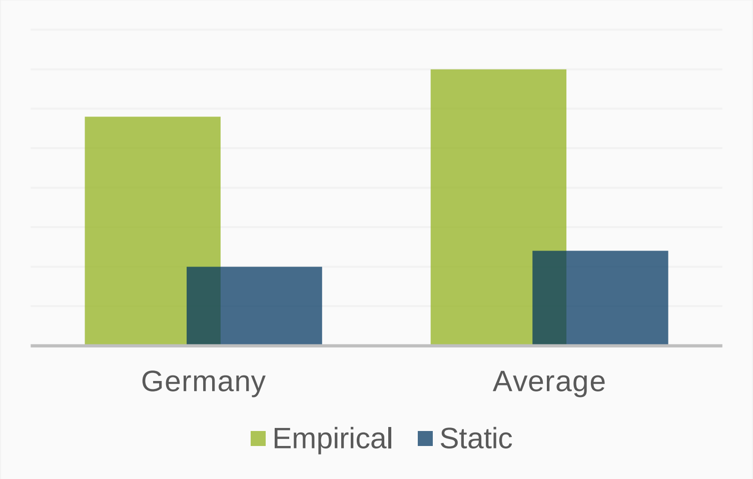
Main Features

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



Evaluation

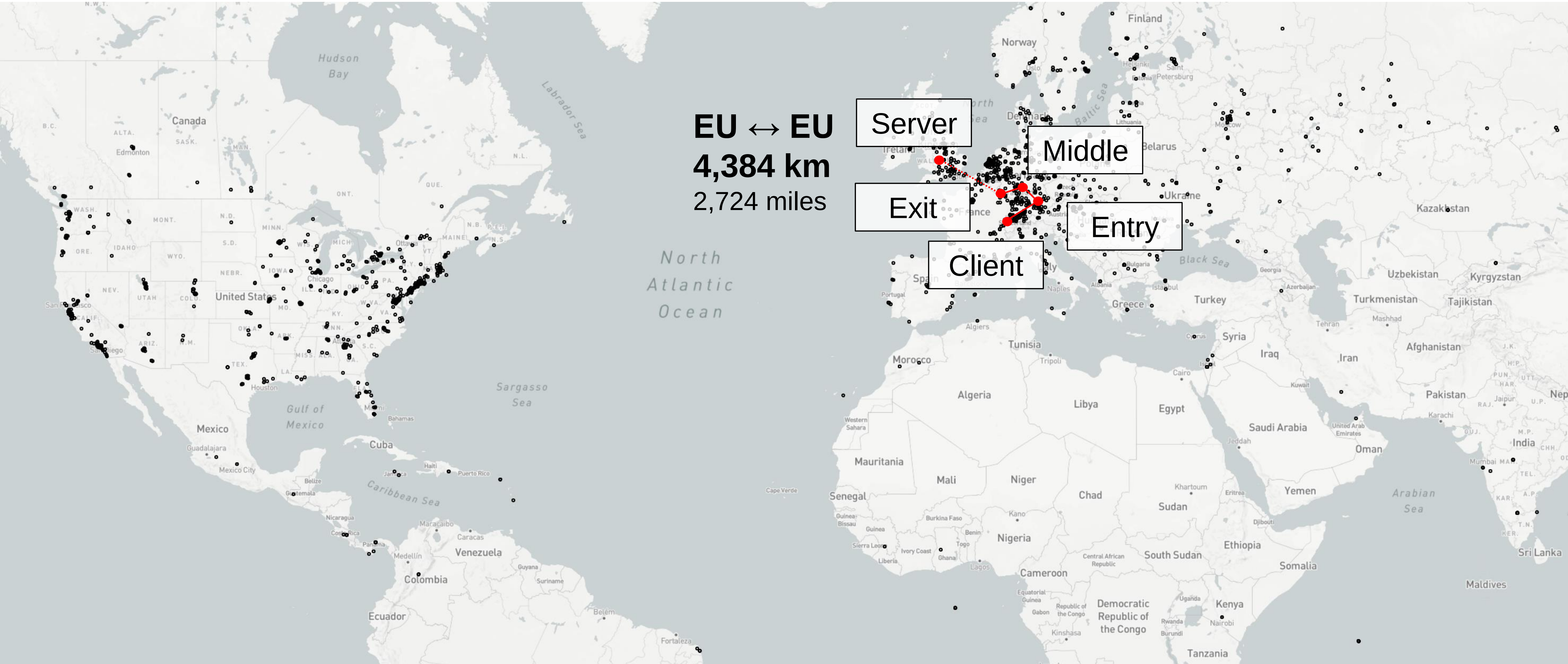
1. Time Ratio for Decision Tradeoff
2. What-if Analysis



Appendix

More information

1. Network Diversity: Connection Lengths

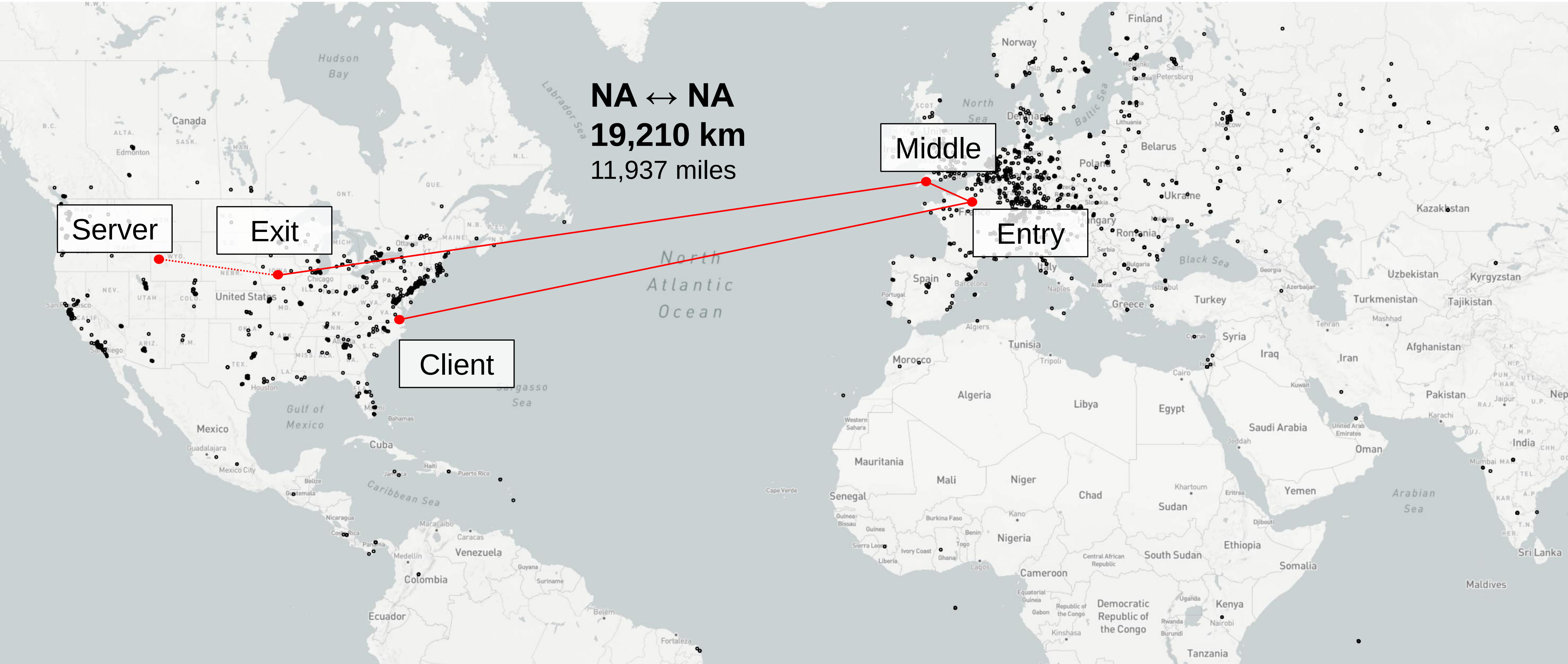


EU ↔ EU

4,384 km

2,724 miles

1. Network Diversity: Connection Lengths



NA ↔ NA

19,210 km

11,937 miles

Server

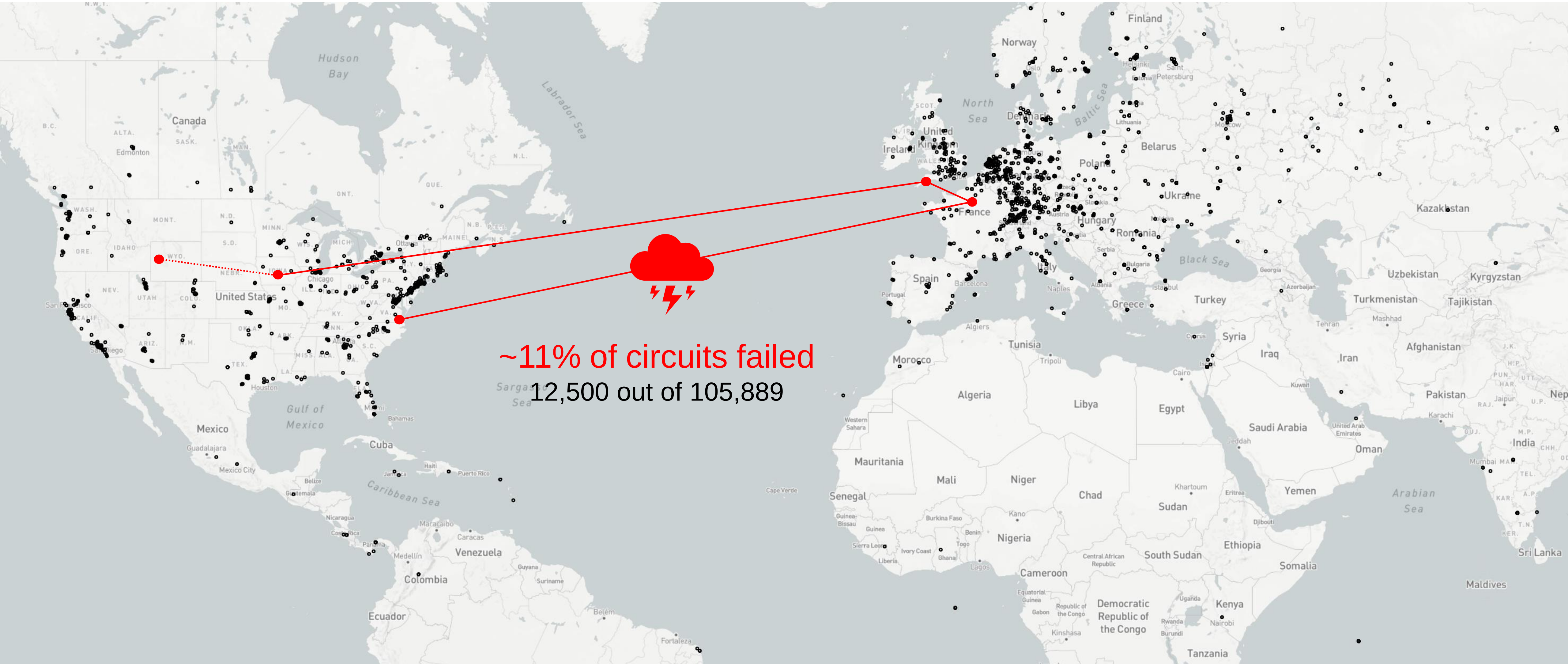
Exit

Middle

Entry

Client

1. Network Diversity: Connection Failures



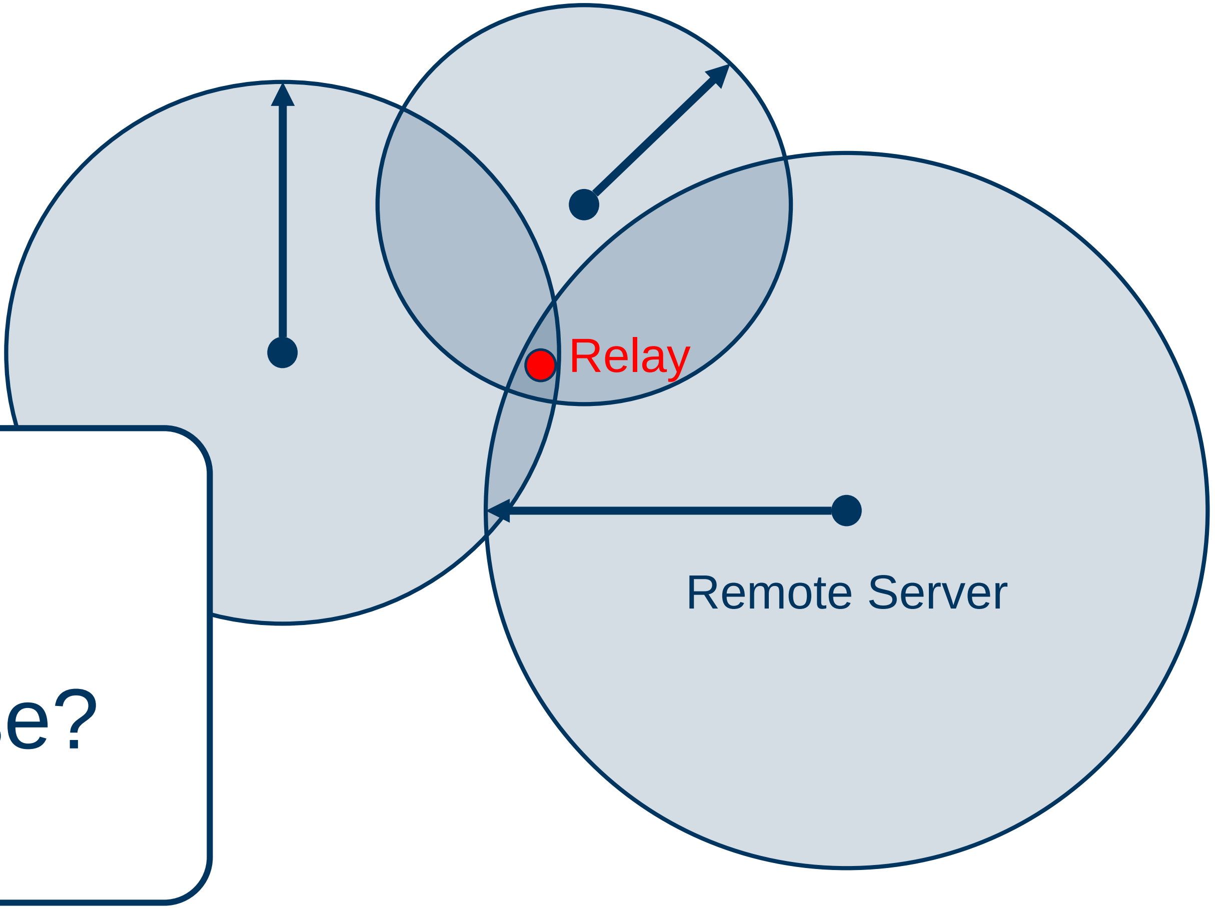
~11% of circuits failed

12,500 out of 105,889

Verification of Relay Locations

2

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

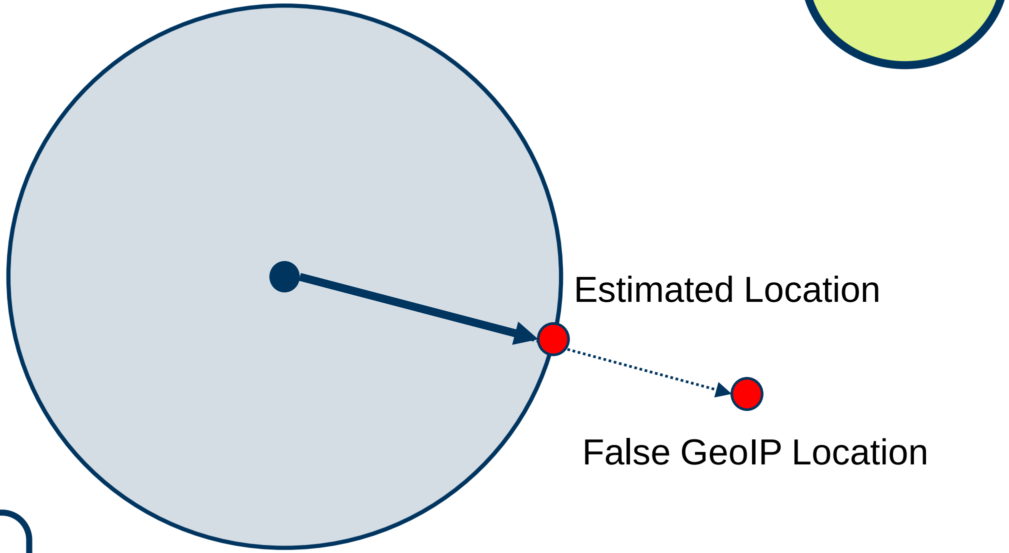


Problem:
Which position is more precise?

Physical Proof

2

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

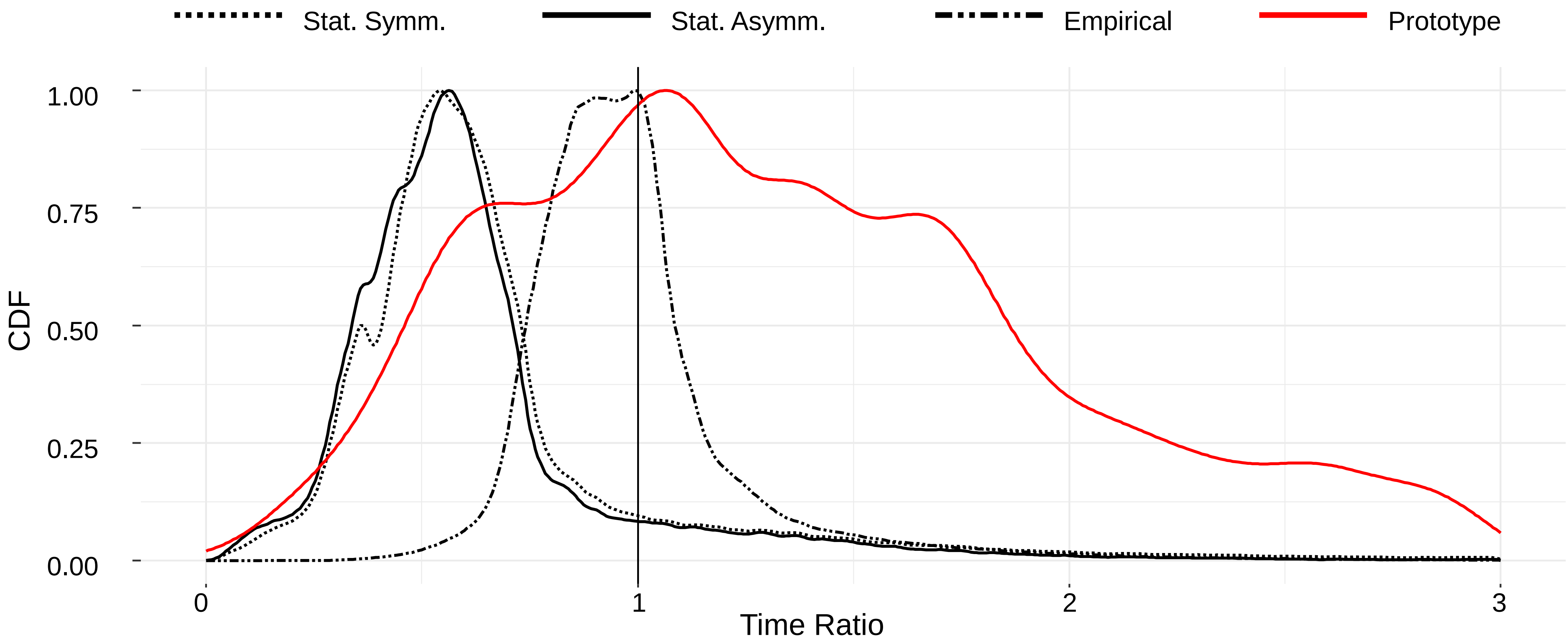


Speed of light proof

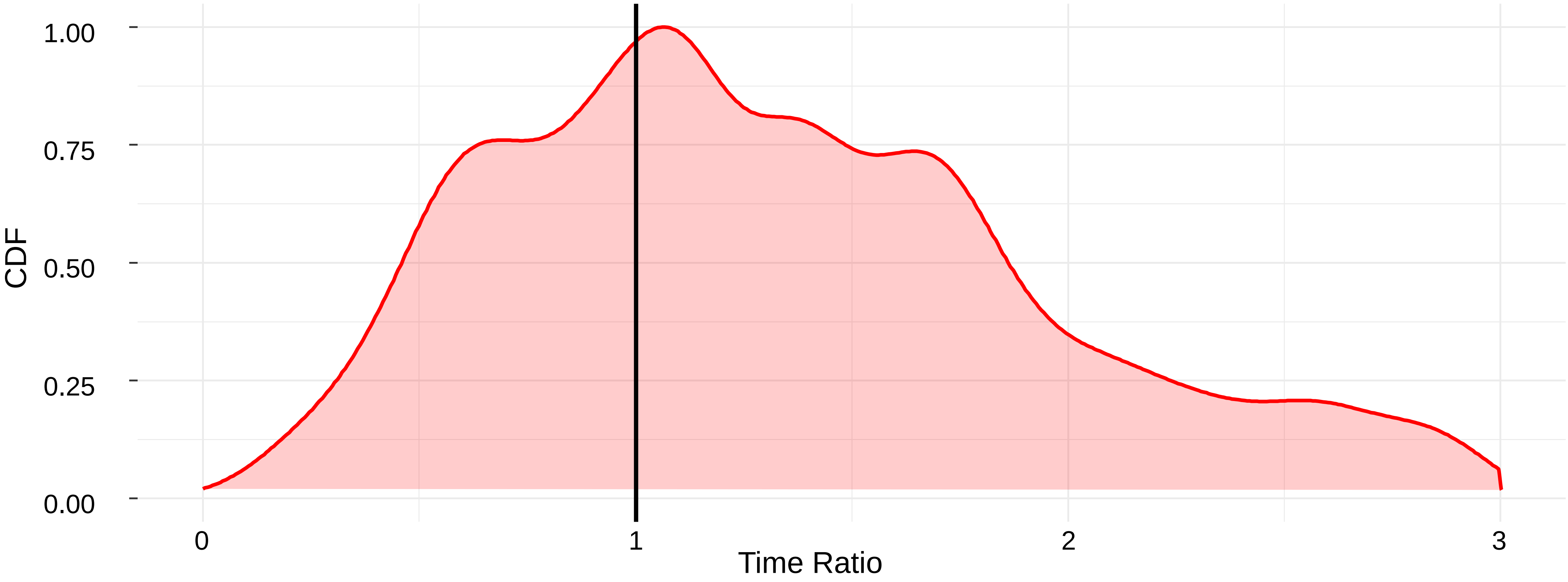
1. Measure RTT from server to relay
2. Compute upper bound threshold with c
   1. Measured Speed ≤ Speed of light?✓
   2. Measured Speed > Speed of light?✗
3. Violation: Update GeoIP location with estimate

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

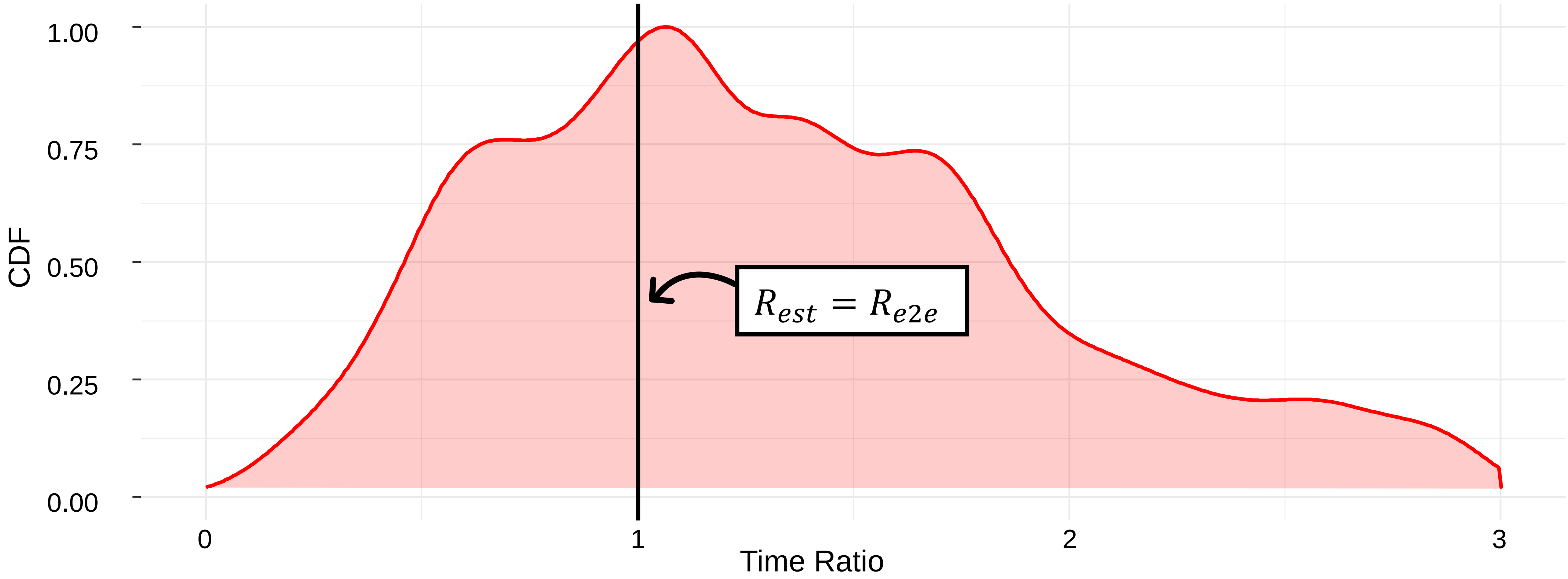


Prototype Simulation



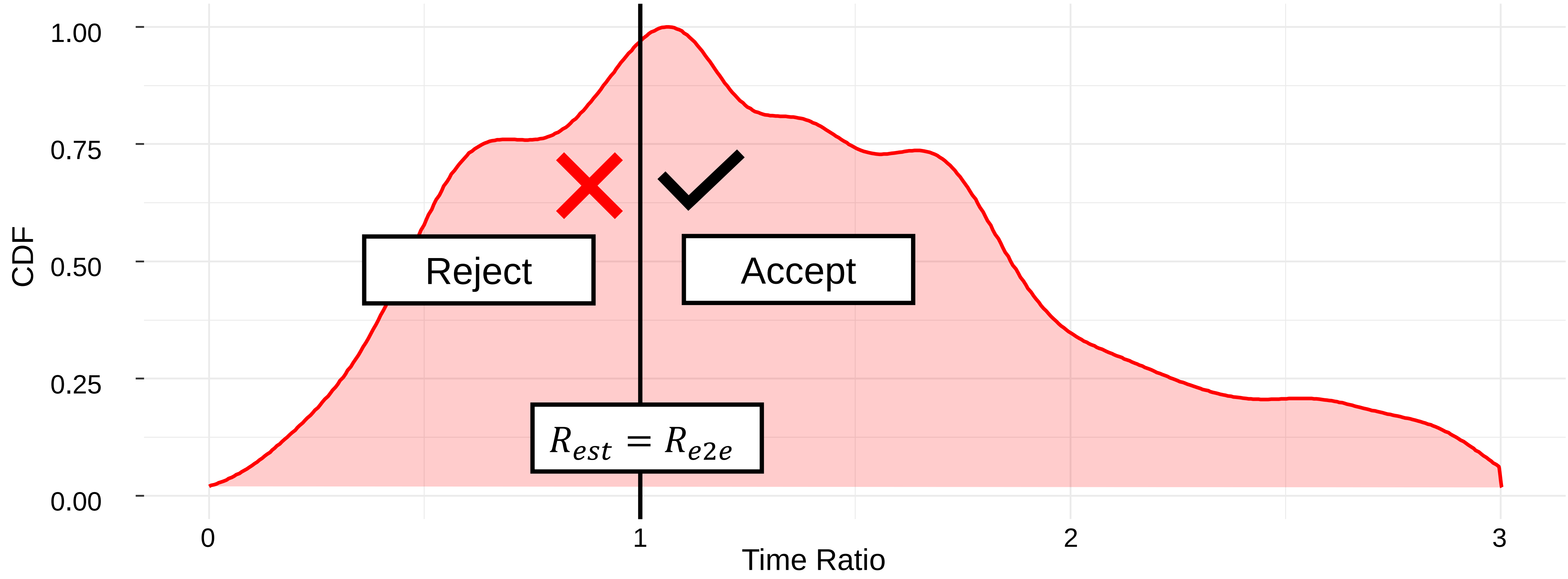
Time Ratio

$$\frac{R_{est}}{R_{e2e}}$$

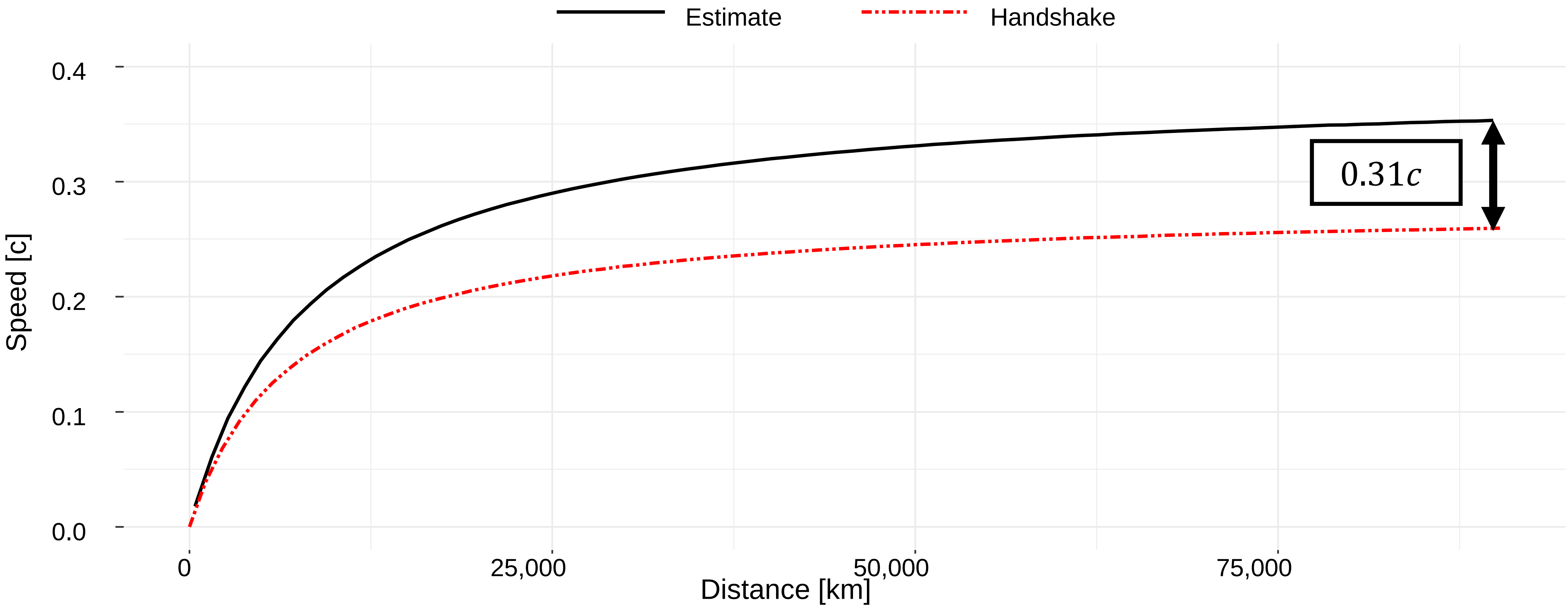


Decision Threshold

$$\frac{R_{est}}{R_{e2e}}$$



Handshake Overhead



Measurement Statistics

Stability of Results

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

Measurement Overhead

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

Experimental Setup

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