Understanding Routing-Induced Censorship Changes Globally

Abhishek Bhaskar, Paul Pearce

Georgia Tech

Censorship Measurement

Georgia Tech

Censorship Measurement

- Internet Censorship is pervasive

Georgia Tech

Censorship Measurement

- Internet Censorship is pervasive
- Significant effort into studying global censorship

Georgia Tech

Censorship Measurement

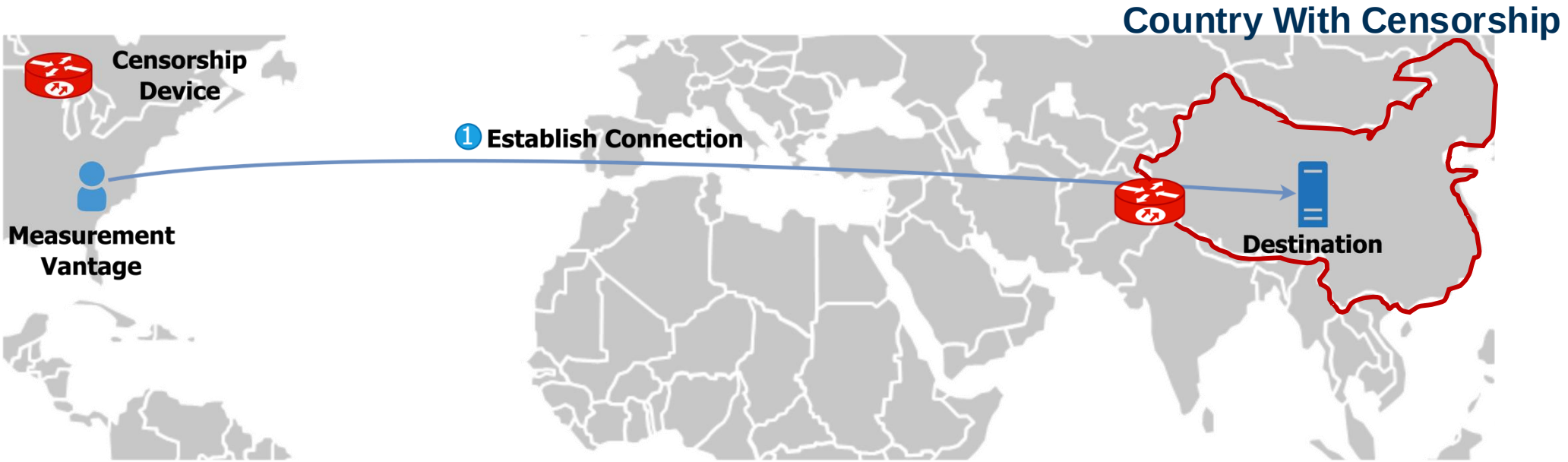
- Internet Censorship is pervasive
- Significant effort into studying global censorship

Outside-In Censorship Measurement

Georgia Tech

Censorship Measurement

- Internet Censorship is pervasive
- Significant effort into studying global censorship

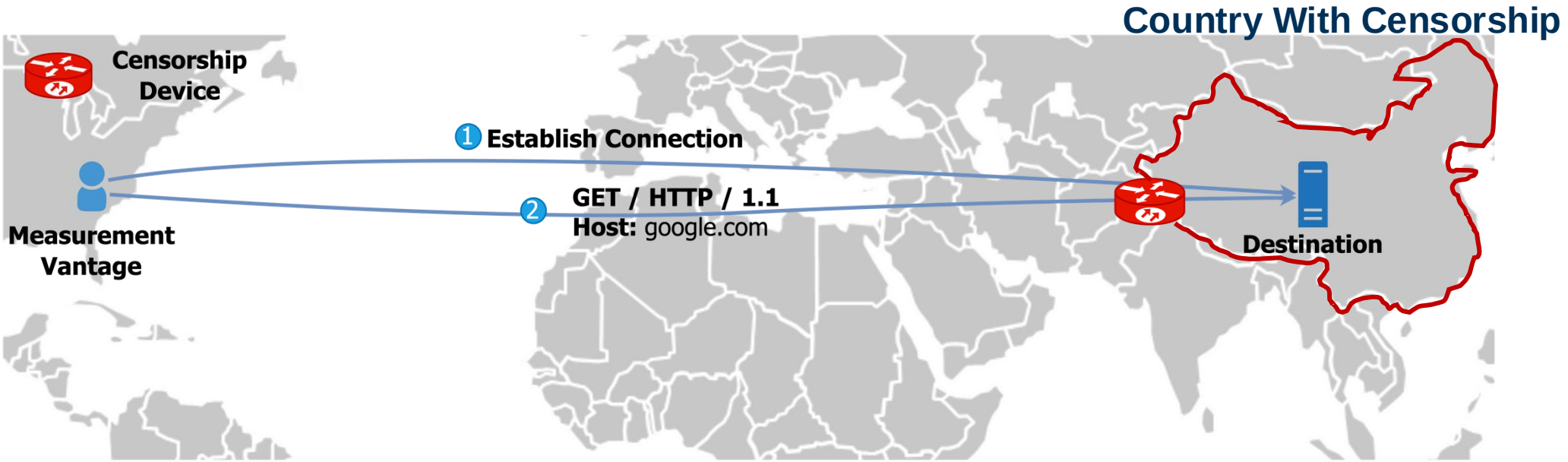


Outside-In Censorship Measurement

Georgia Tech

Censorship Measurement

- Internet Censorship is pervasive
- Significant effort into studying global censorship

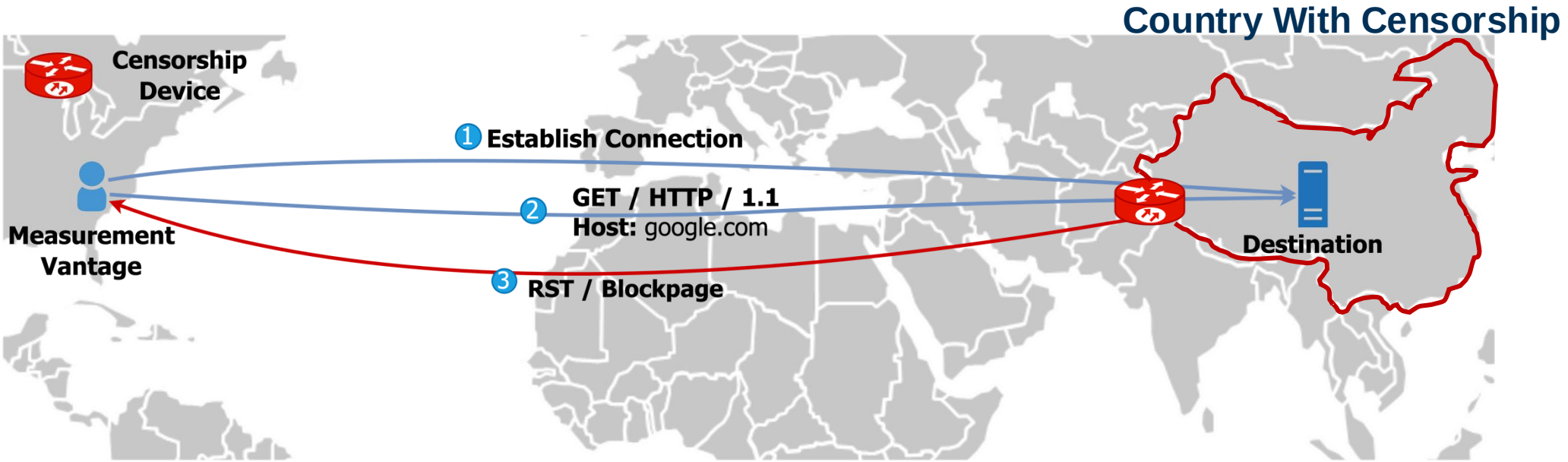


Outside-In Censorship Measurement

Georgia Tech

Censorship Measurement

- Internet Censorship is pervasive
- Significant effort into studying global censorship



Outside-In Censorship Measurement

Georgia Tech

Equal-Cost Multi-Path (ECMP) Routing

Georgia Tech

Equal-Cost Multi-Path (ECMP) Routing

- Routers use parts of the packet (flow-id [1]) to load balance

[1] Augustin, Brice, et al. **"Avoiding traceroute anomalies with Paris traceroute."**

Georgia Tech

Equal-Cost Multi-Path (ECMP) Routing

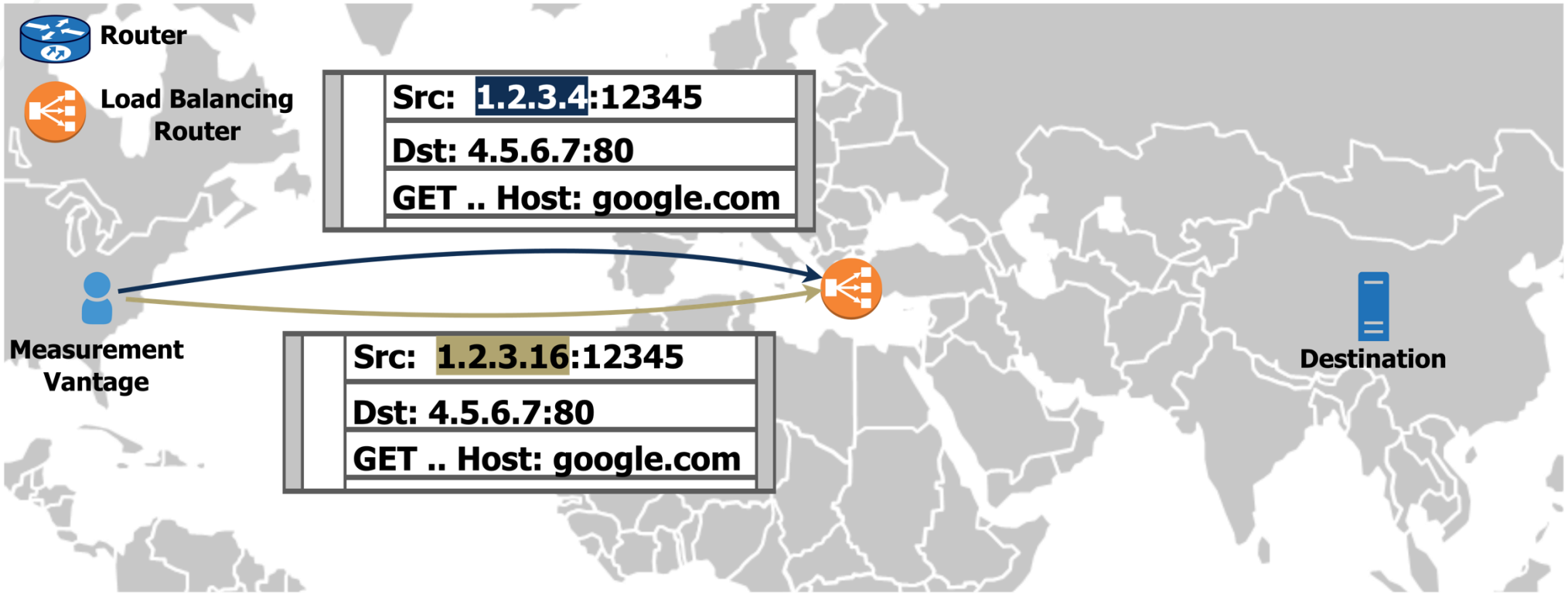
- Routers use parts of the packet (flow-id [1]) to load balance

[1] Augustin, Brice, et al. **"Avoiding traceroute anomalies with Paris traceroute."**

Georgia Tech

Equal-Cost Multi-Path (ECMP) Routing

- Routers use parts of the packet (flow-id [1]) to load balance

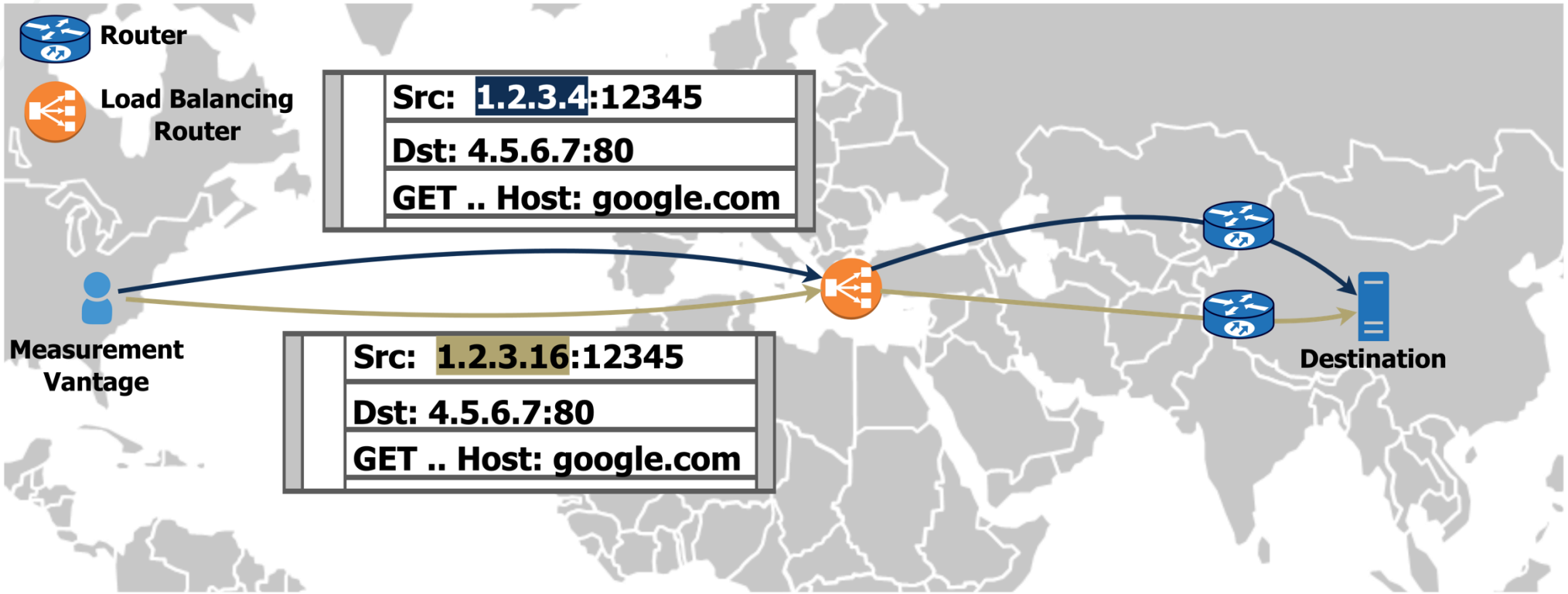


[1] Augustin, Brice, et al. **"Avoiding traceroute anomalies with Paris traceroute."**

Georgia Tech

Equal-Cost Multi-Path (ECMP) Routing

- Routers use parts of the packet (flow-id [1]) to load balance

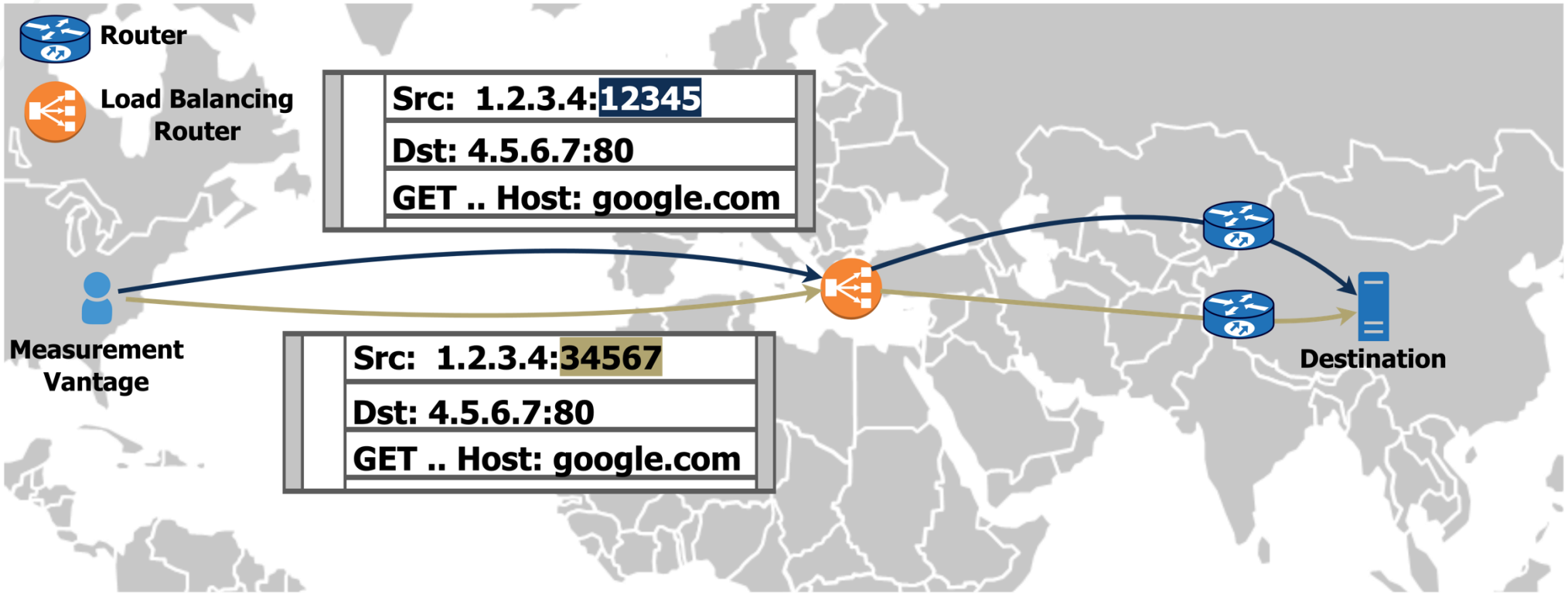


[1] Augustin, Brice, et al. **"Avoiding traceroute anomalies with Paris traceroute."**

Georgia Tech

Equal-Cost Multi-Path (ECMP) Routing

- Routers use parts of the packet (flow-id [1]) to load balance

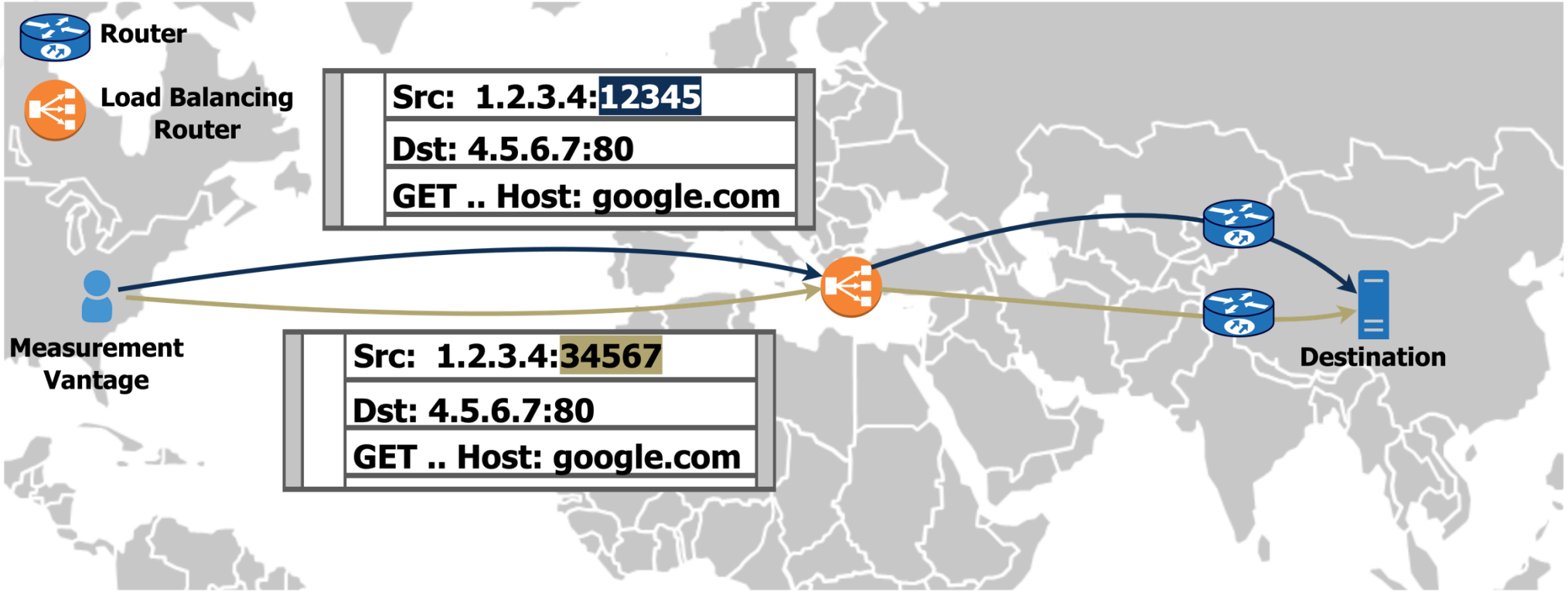


[1] Augustin, Brice, et al. **"Avoiding traceroute anomalies with Paris traceroute."**

Georgia Tech

Equal-Cost Multi-Path (ECMP) Routing

- Routers use parts of the packet (flow-id [1]) to load balance



- **Diamond-Miner** [2] found that **18% of divergences span ASes**

[1] Augustin, Brice, et al. **"Avoiding traceroute anomalies with Paris traceroute."**
[2] Vermeulen, Kevin, et al. "**Diamond-Miner: Comprehensive Discovery of the Internet's Topology Diamonds.**"

Georgia Tech

Variance in Censorship Measurement

- Prior works have observed non-determinism in censorship results

Georgia Tech

Variance in Censorship Measurement

- Prior works have observed non-determinism in censorship results

> ”One set of domains in China experiences manipulation by approximately 80% of resolvers, and another set **experiences manipulation only half the time**"
>
> - Pearce, Paul, et al. "Global measurement of DNS manipulation."

Georgia Tech

Variance in Censorship Measurement

- Prior works have observed non-determinism in censorship results

”One set of domains in China experiences manipulation by approximately 80% of resolvers, and another set **experiences manipulation only half the time**"

- Pearce, Paul, et al. "Global measurement of DNS manipulation."

"Furthermore, in May 2016 and May 2017, we have observed that RST packets sometimes were **unable to tear down the TCB** on the GFW, with **different pairs of controlled clients and servers**."

- Wang, Zhongjie, et al. "Your state is not mine: A closer look at evading stateful internet censorship."

Georgia Tech

Variance in Censorship Measurement

- Prior works have observed non-determinism in censorship results

”One set of domains in China experiences manipulation by approximately 80% of resolvers, and another set **experiences manipulation only half the time**"

- Pearce, Paul, et al. "Global measurement of DNS manipulation."

"Furthermore, in May 2016 and May 2017, we have observed that RST packets sometimes were **unable to tear down the TCB** on the GFW, with **different pairs of controlled clients and servers**."

- Wang, Zhongjie, et al. "Your state is not mine: A closer look at evading stateful internet censorship."

"ISP and organizational blockpages have **infrequent absences…”**

- Raman, Ram Sundara, et al. "Measuring the Deployment of Network Censorship Filters at Global Scale."

Georgia Tech

Variance in Censorship Measurement

- Prior works have observed non-determinism in censorship results

”One set of domains in China experiences manipulation by approximately 80% of resolvers, and another set **experiences manipulation only half the time**"

- Pearce, Paul, et al. "Global measurement of DNS manipulation."

"ISP and organizational blockpages have **infrequent absences…”**

- Raman, Ram Sundara, et al. "Measuring the Deployment of Network Censorship Filters at Global Scale."

"Furthermore, in May 2016 and May 2017, we have observed that RST packets sometimes were **unable to tear down the TCB** on the GFW, with **different pairs of controlled clients and servers**."

- Wang, Zhongjie, et al. "Your state is not mine: A closer look at evading stateful internet censorship."

“We find that **different ISPs are blocking different websites** in India, despite ostensibly being served the same legal orders.”

- Katira, Divyank, et al. "CensorWatch: On the Implementation of Online Censorship in India."

"**We observe considerable variation**, across countries, over time, and across ISPs in the same country, or even across different types of content that were tested in the same ISP"

- Gill, Phillipa, et al. "Characterizing web censorship worldwide: Another look at the opennet initiative data."

Georgia Tech

Variance in Censorship Measurement

- Prior works have observed non-determinism in censorship results

”One set of domains in China experiences manipulation by approximately 80% of resolvers, and another set **experiences manipulation only half the time**"

- Pearce, Paul, et al. "Global measurement of DNS manipulation."

"Furthermore, in May 2016 and May 2017, we have observed that RST packets sometimes were **unable to tear down the TCB** on the GFW, with **different pairs of controlled clients and servers**."

- Wang, Zhongjie, et al. "Your state is not mine: A closer look at evading stateful internet censorship."

“We find that **different ISPs are blocking different websites** in India, despite ostensibly being served the same legal orders.”

- Katira, Divyank, et al. "CensorWatch: On the Implementation of Online Censorship in India."

"ISP and organizational blockpages have **infrequent absences…”**

- Raman, Ram Sundara, et al. "Measuring the Deployment of Network Censorship Filters at Global Scale."

"**We observe considerable variation**, across countries, over time, and across ISPs in the same country, or even across different types of content that were tested in the same ISP"

- Gill, Phillipa, et al. "Characterizing web censorship worldwide: Another look at the opennet initiative data."

- **Important to rule out measurement artifacts as potential cause for variation**

Georgia Tech

Our Work

Georgia Tech

Our Work

- Study the impact of ECMP routing on **global outside-in censorship measurement**

Georgia Tech

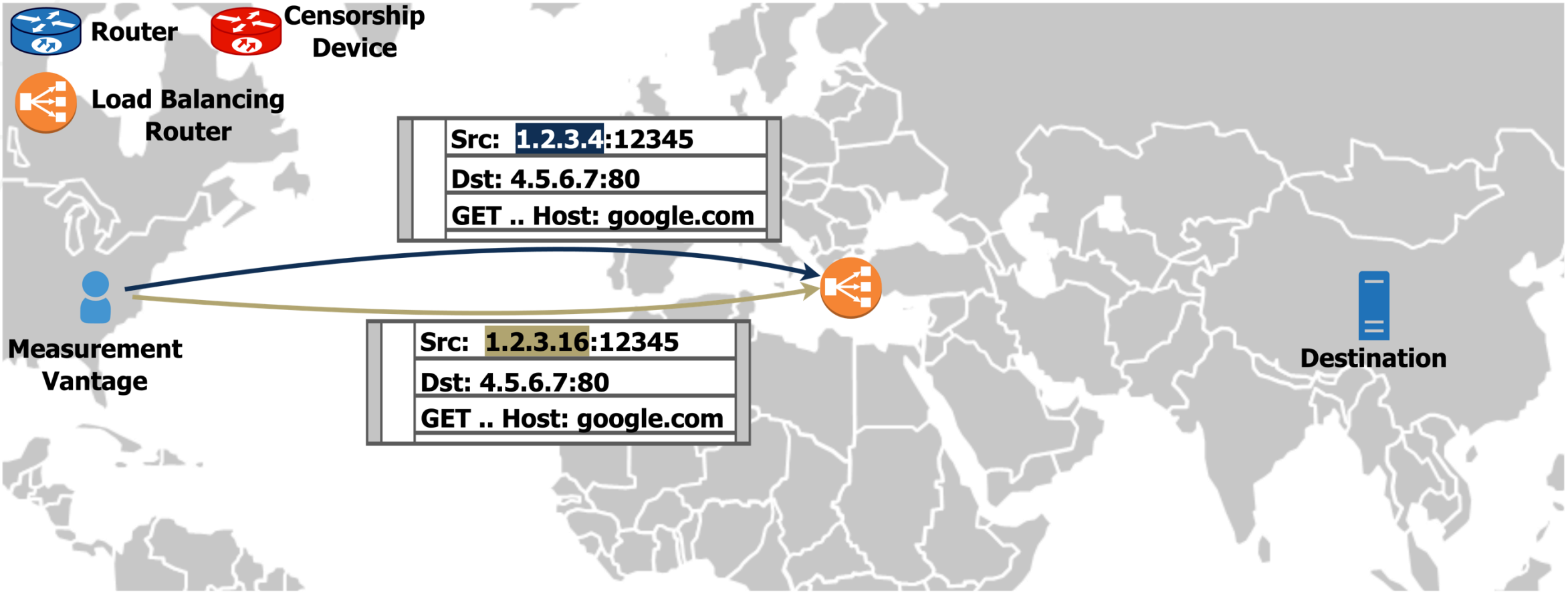
Our Work

- Study the impact of ECMP routing on **global outside-in censorship measurement**

Georgia Tech

Our Work

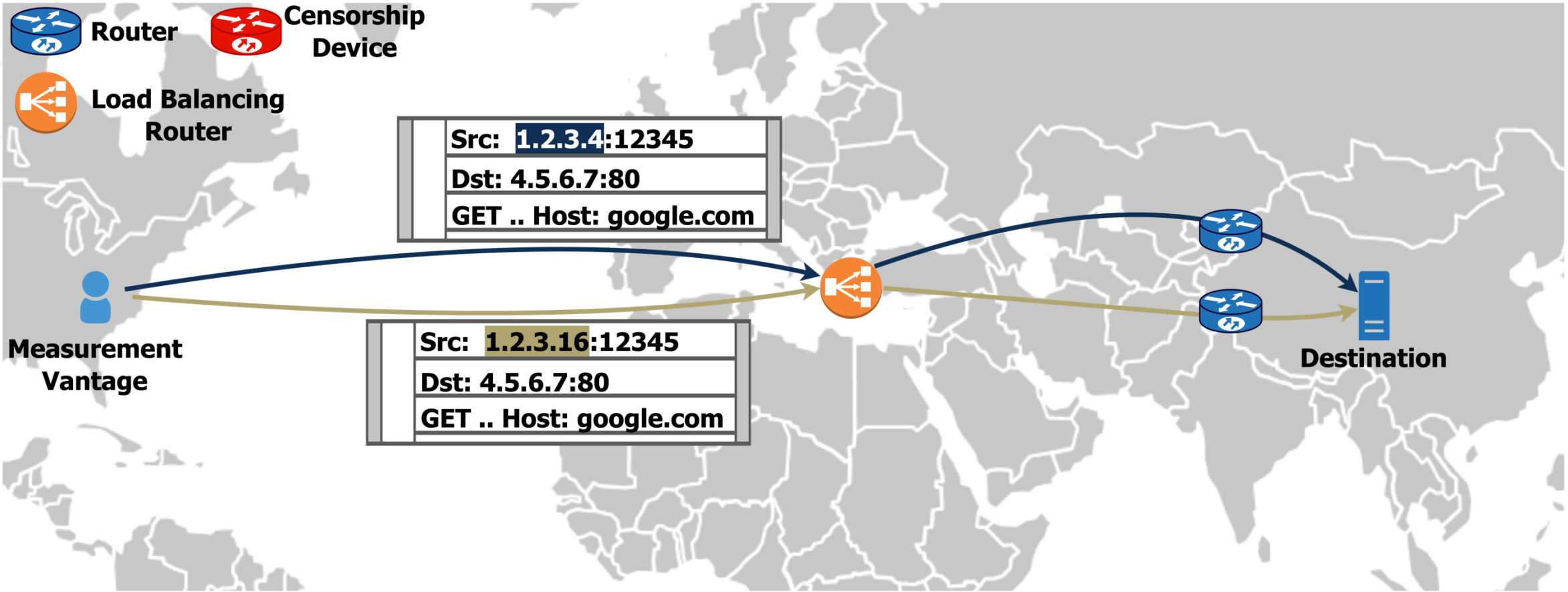
- Study the impact of ECMP routing on **global outside-in censorship measurement**



Georgia Tech

Our Work

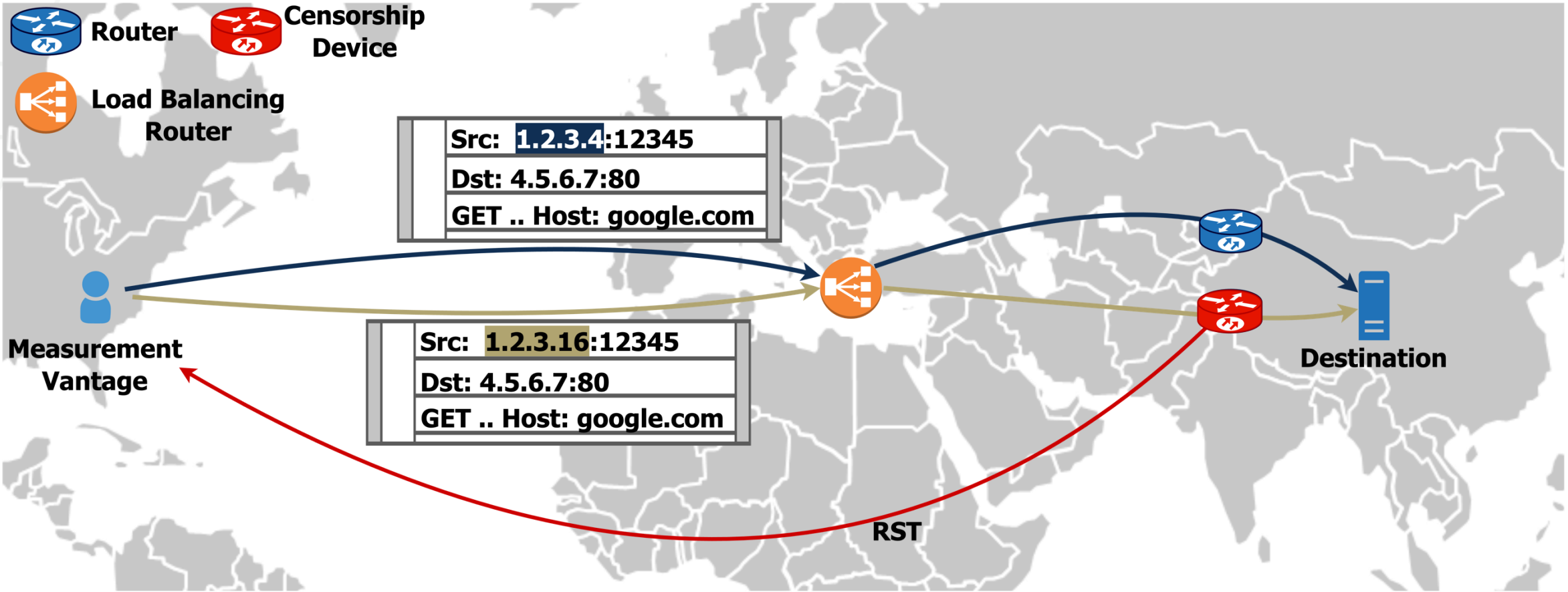
- Study the impact of ECMP routing on **global outside-in censorship measurement**



Georgia Tech

Our Work

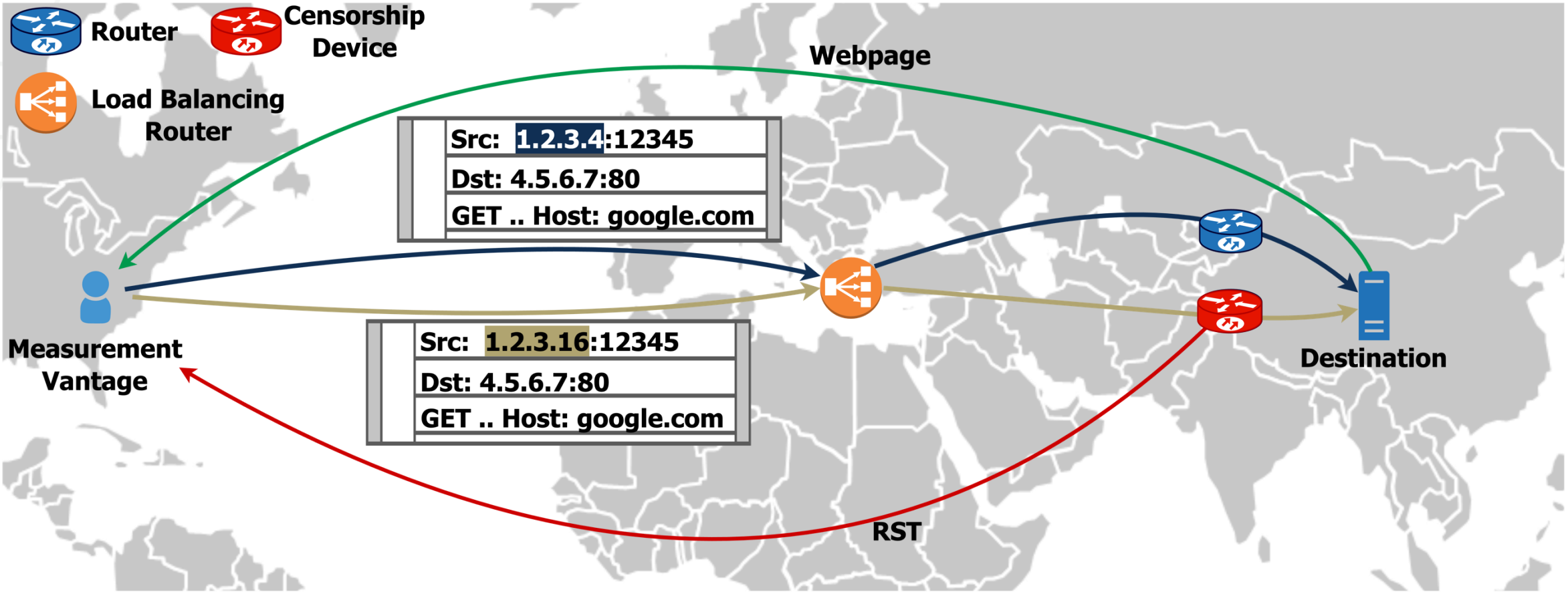
- Study the impact of ECMP routing on **global outside-in censorship measurement**



Georgia Tech

Our Work

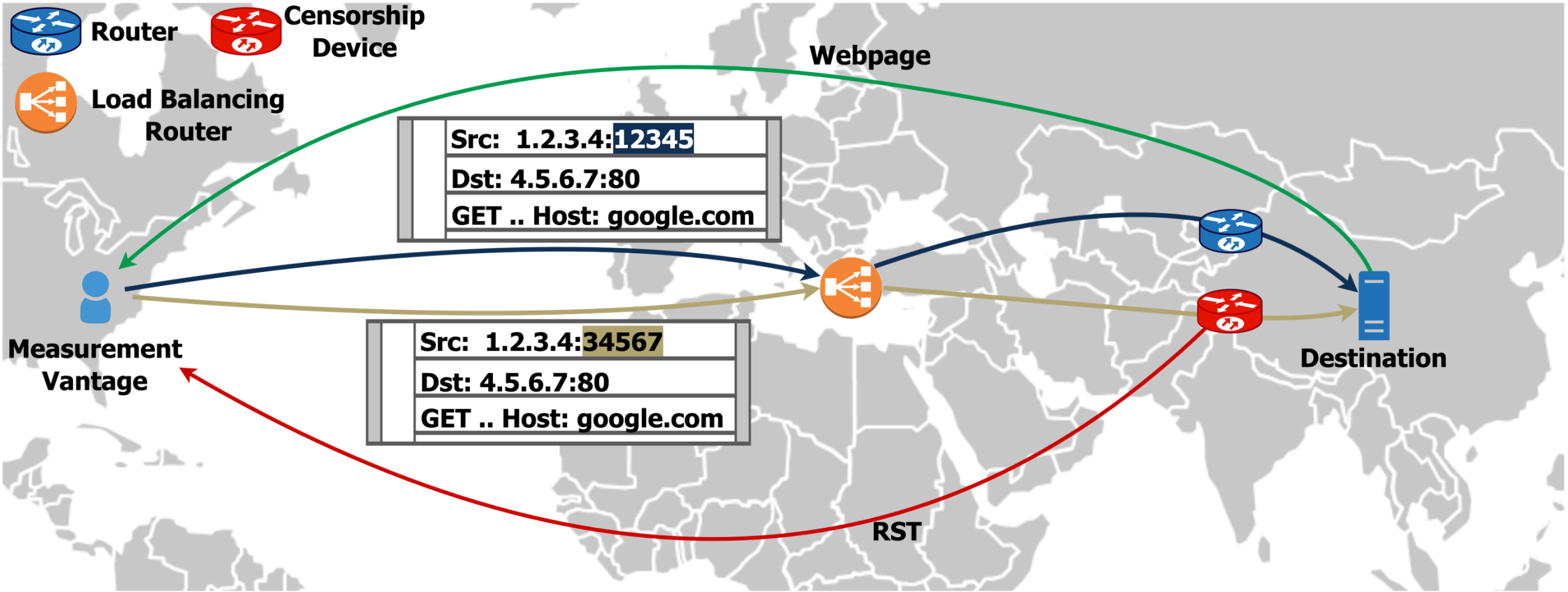
- Study the impact of ECMP routing on **global outside-in censorship measurement**



Georgia Tech

Our Work

- Study the impact of ECMP routing on **global outside-in censorship measurement**

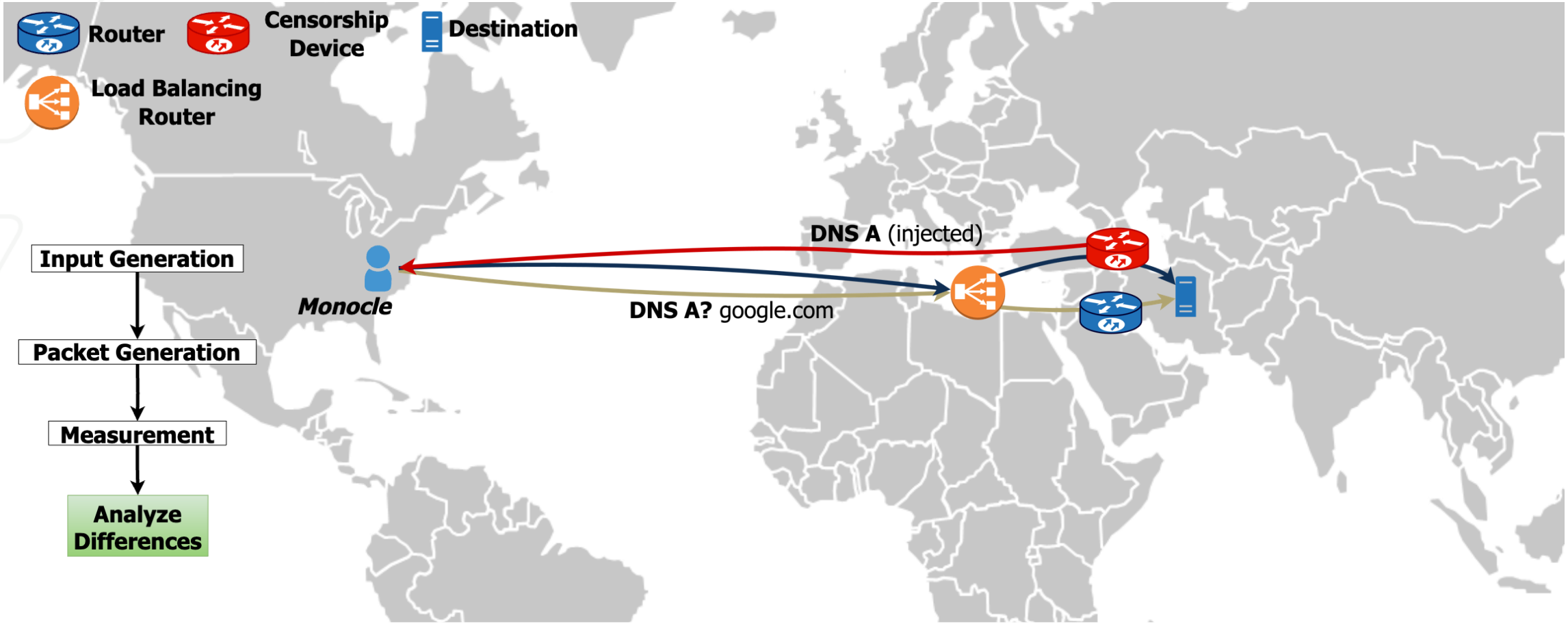


Georgia Tech

Monocle

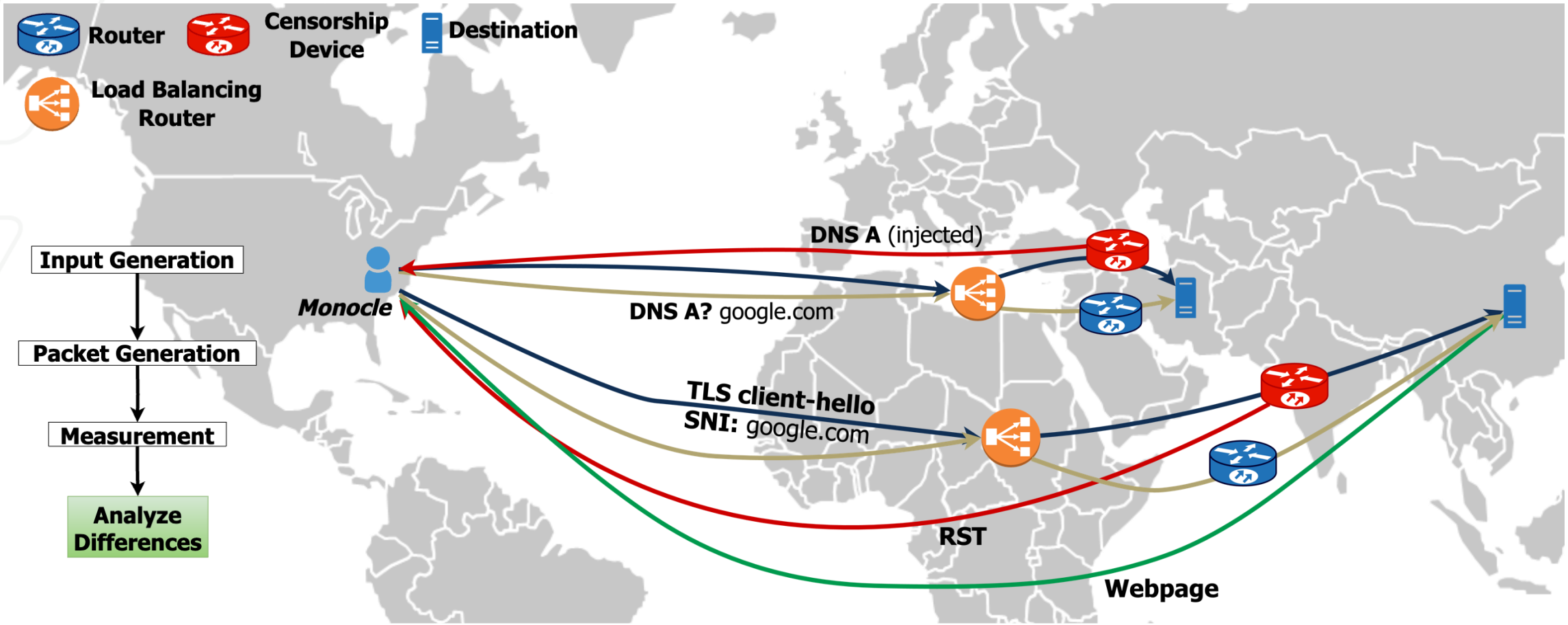
Georgia Tech

Monocle



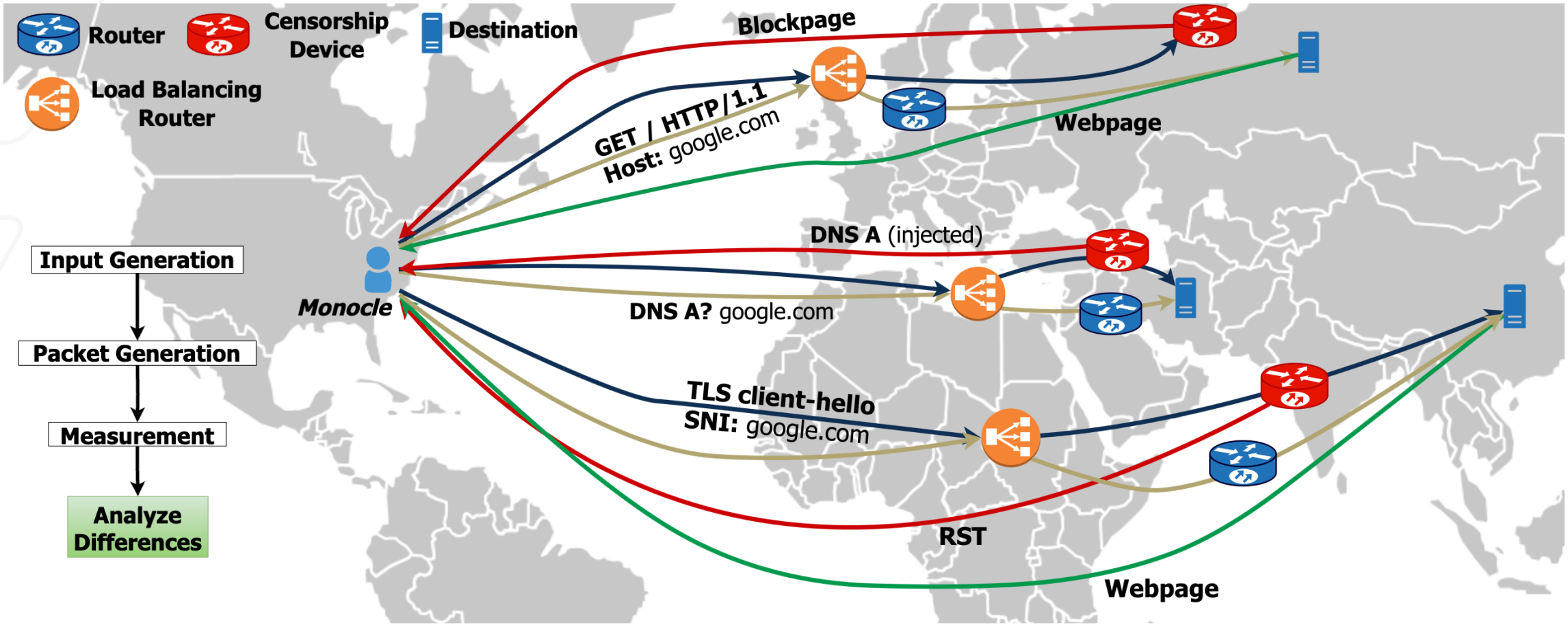
Georgia Tech

Monocle



Georgia Tech

Monocle



Georgia Tech

Route-Aware Traceroute

Georgia Tech

Route-Aware Traceroute

- To understand how ECMP causes variation

Georgia Tech

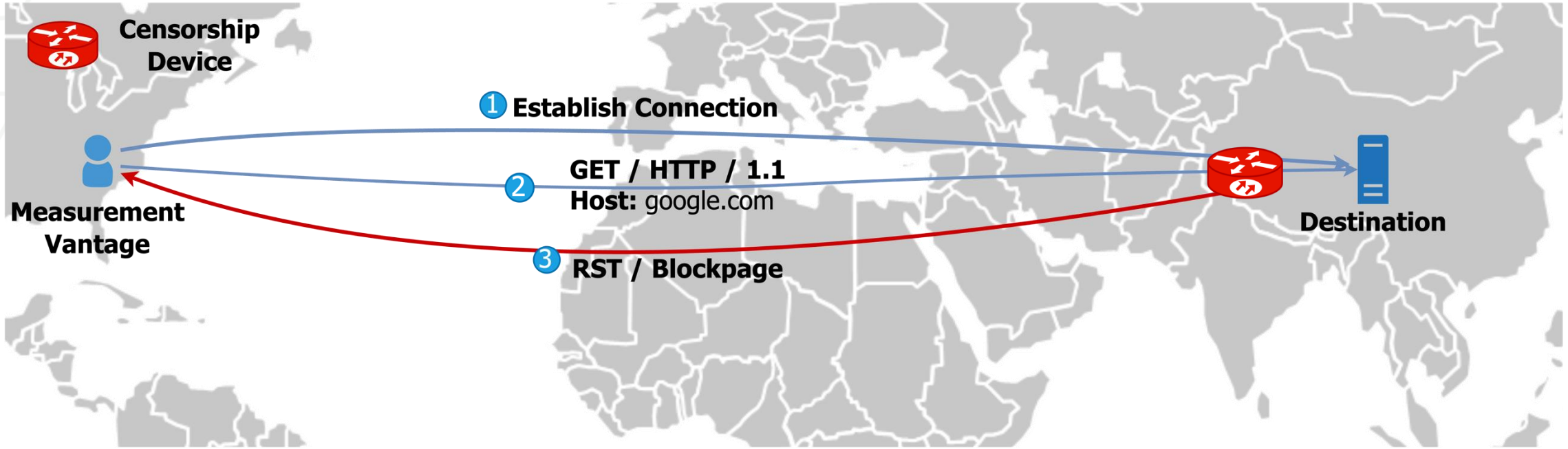
Route-Aware Traceroute

- To understand how ECMP causes variation
  - We build route-stable network paths (UDP and TCP)

Georgia Tech

Route-Aware Traceroute

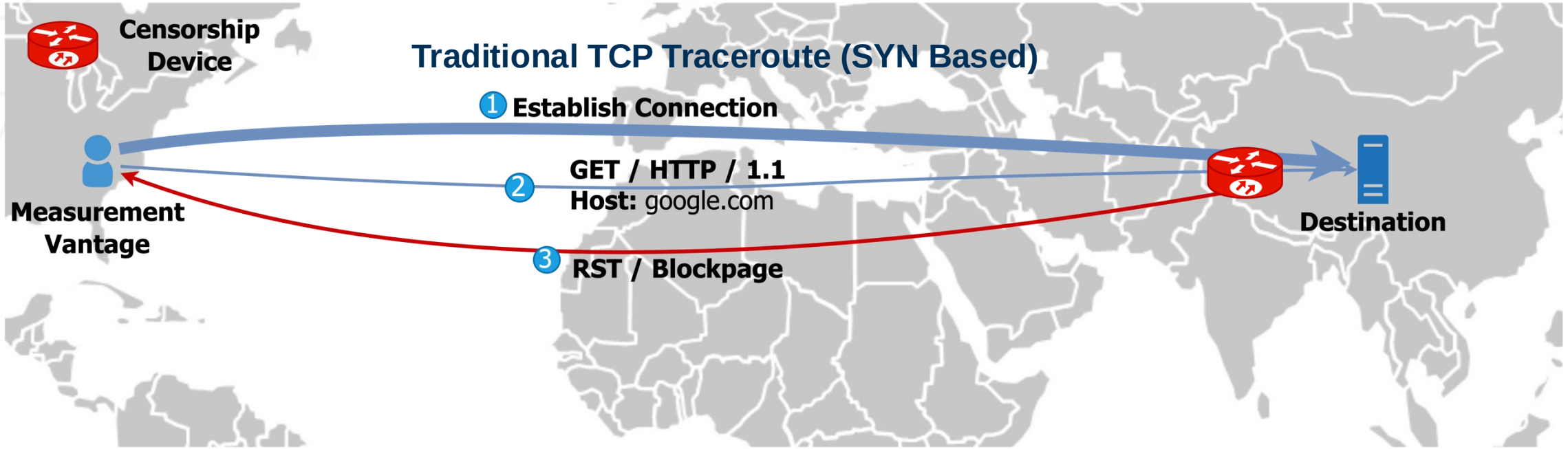
- To understand how ECMP causes variation
  - We build route-stable network paths (UDP and TCP)



Georgia Tech

Route-Aware Traceroute

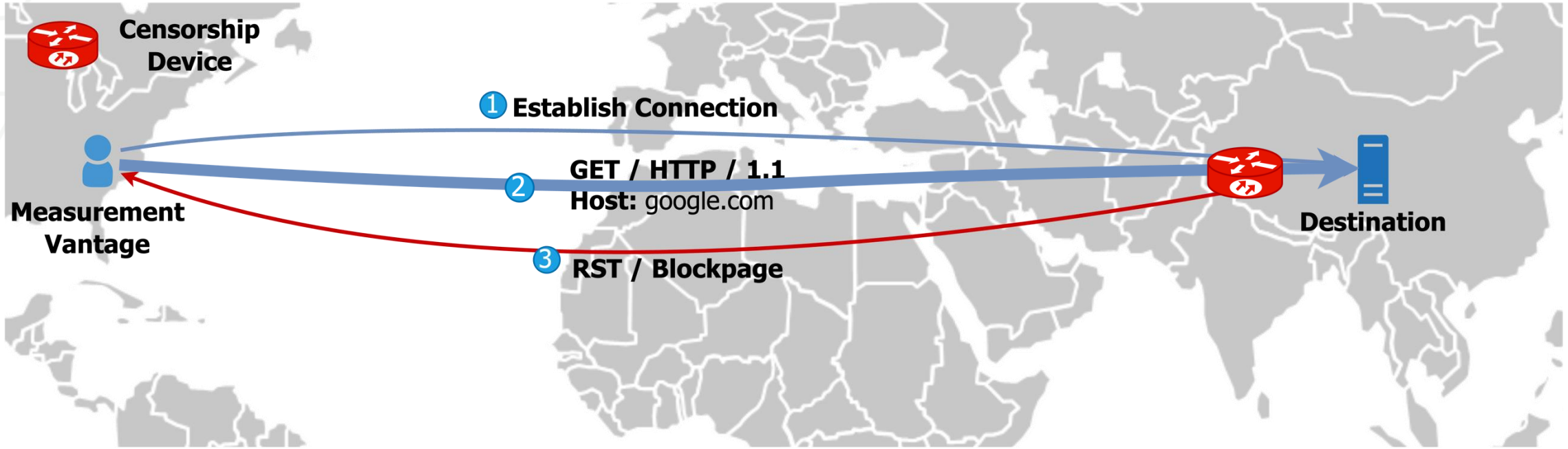
- To understand how ECMP causes variation
  - We build route-stable network paths (UDP and TCP)



Georgia Tech

Route-Aware Traceroute

- To understand how ECMP causes variation
  - We build route-stable network paths (UDP and TCP)



- Challenging as sensitive packet is **deep inside TCP connection**

Georgia Tech

Questions we want to answer

Georgia Tech

Questions we want to answer

- What is the **path diversity** of global censorship measurement?

Georgia Tech

Questions we want to answer

- What is the **path diversity** of global censorship measurement?
- High-level impact of ECMP on **global censorship measurement results**?

Georgia Tech

Questions we want to answer

- What is the **path diversity** of global censorship measurement?
- High-level impact of ECMP on **global censorship measurement results**?
- How do **changes in the packet** influence censorship results?

Georgia Tech

Questions we want to answer

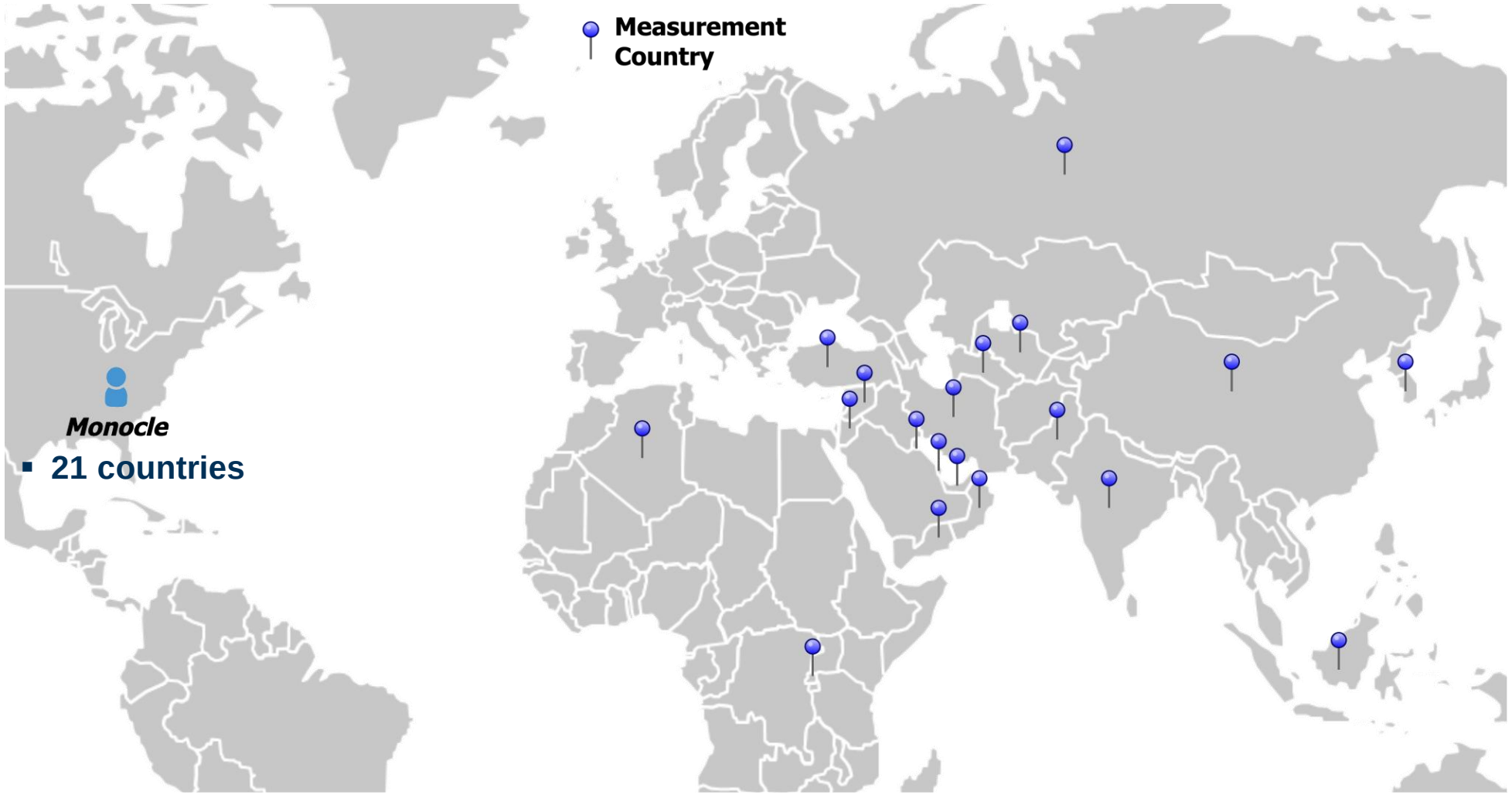
- What is the **path diversity** of global censorship measurement?
- High-level impact of ECMP on **global censorship measurement results**?
- How do **changes in the packet** influence censorship results?
- What **underlying network structures** cause change?

Questions we want to answer

- What is the **path diversity** of global censorship measurement?
- High-level impact of ECMP on **global censorship measurement results**?
- How do **changes in the packet** influence censorship results?
- What **underlying network structures** cause change?
- How do results contextualize with prior work?

Georgia Tech

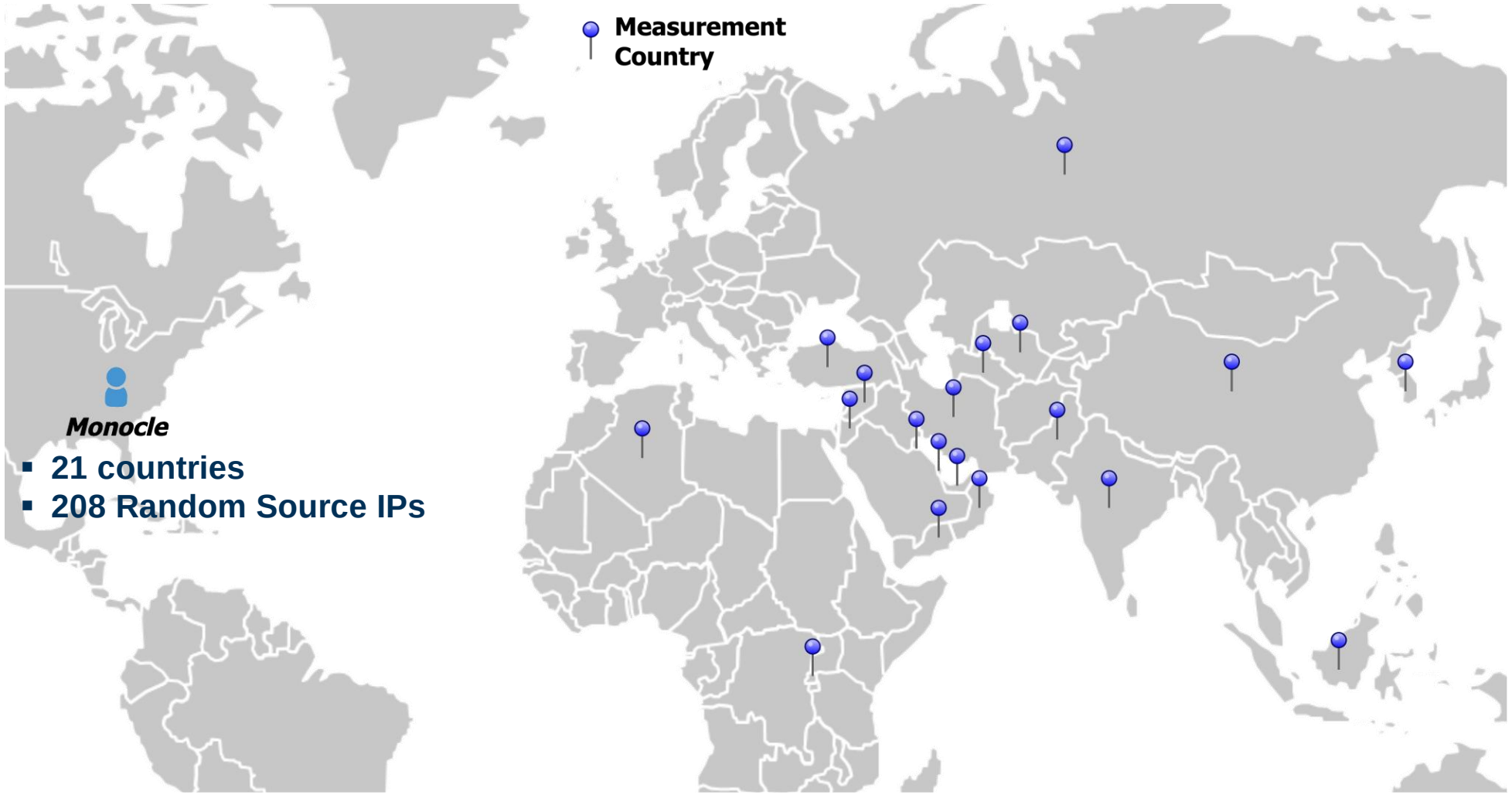
Dataset



- 21 countries

Georgia Tech

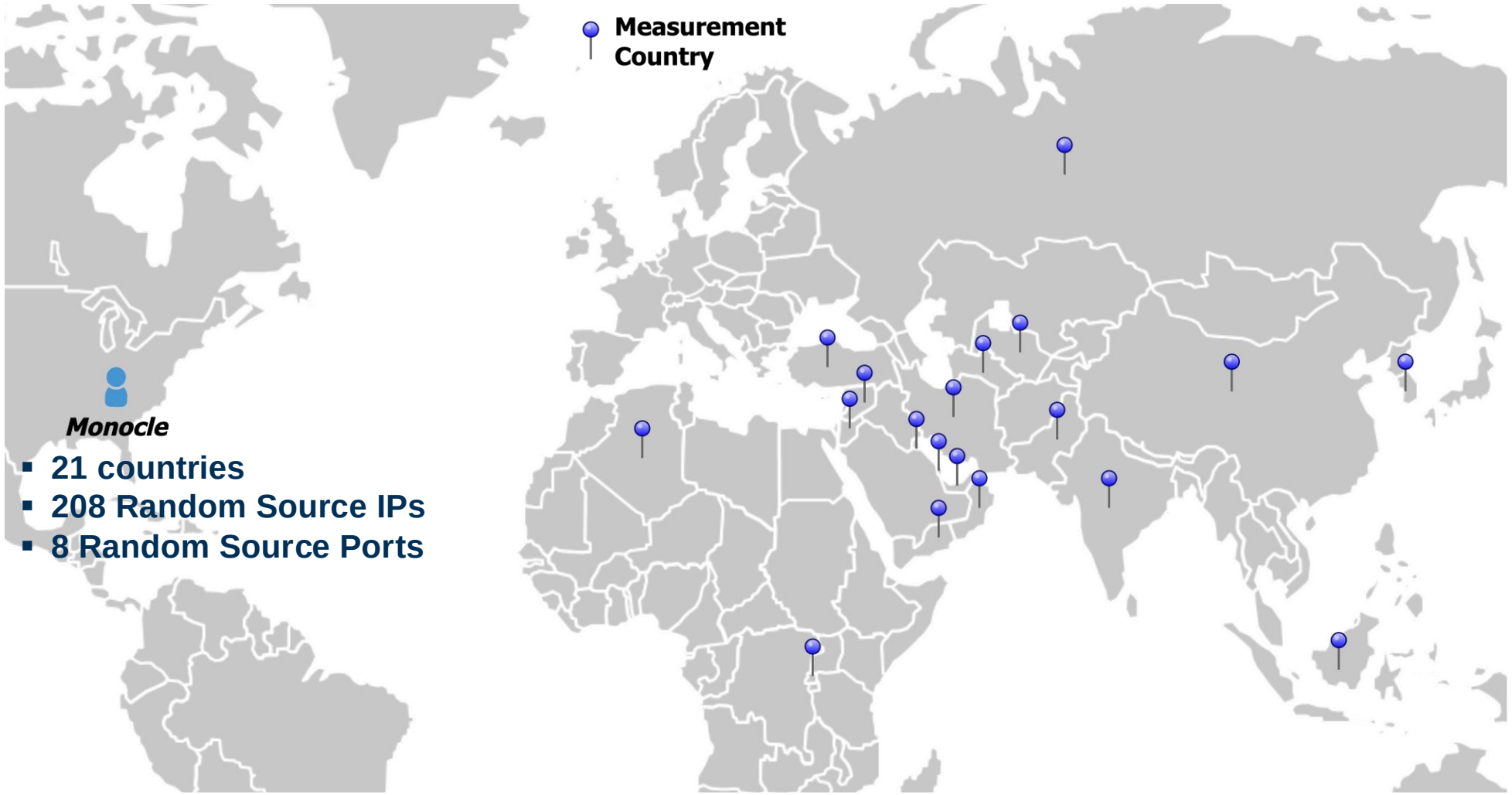
Dataset



- 21 countries
- 208 Random Source IPs

Georgia Tech

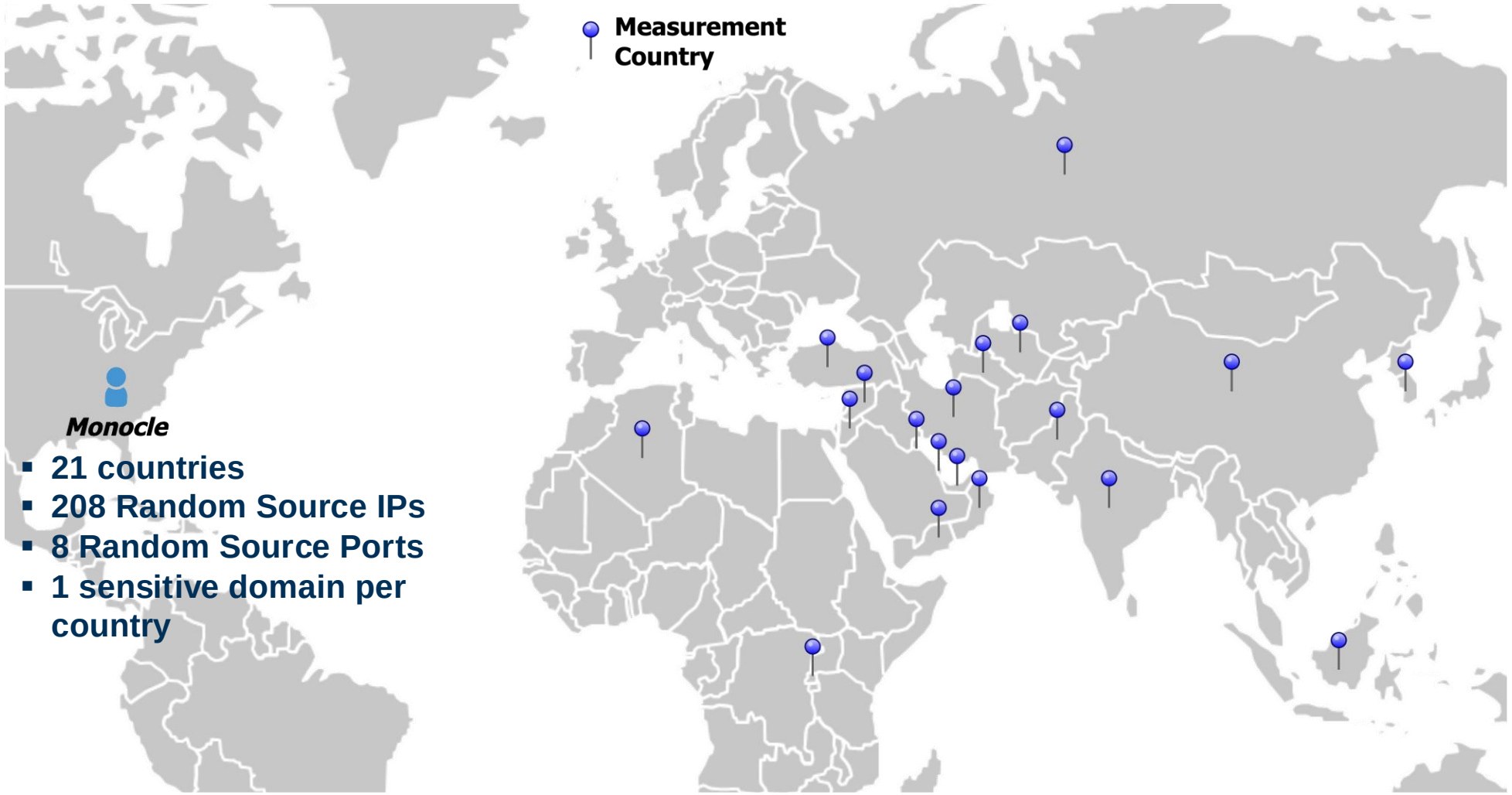
Dataset



- 21 countries
- 208 Random Source IPs
- 8 Random Source Ports

Georgia Tech

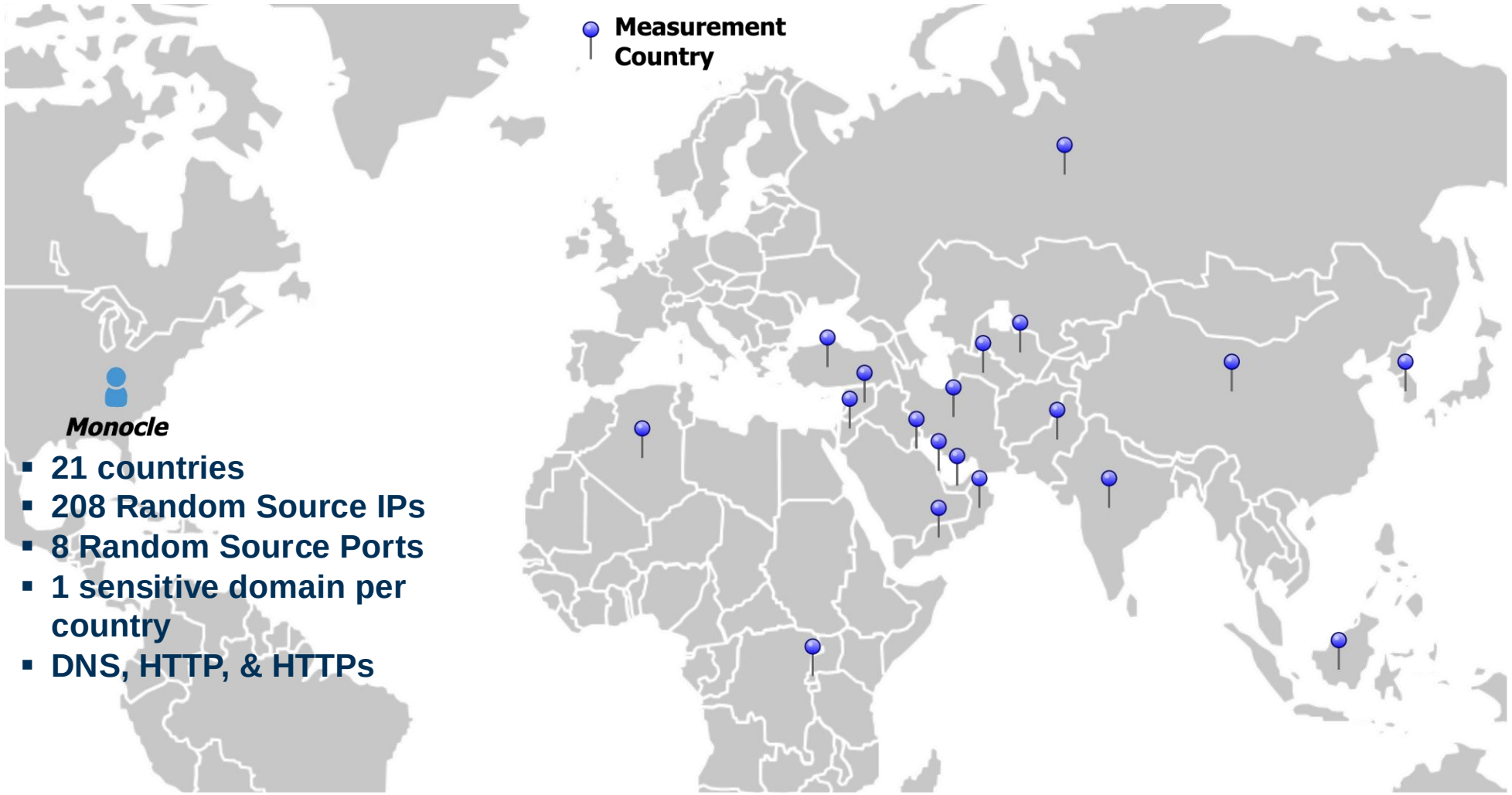
Dataset



- 21 countries
- 208 Random Source IPs
- 8 Random Source Ports
- 1 sensitive domain per country

Georgia Tech

Dataset



- 21 countries
- 208 Random Source IPs
- 8 Random Source Ports
- 1 sensitive domain per country
- DNS, HTTP, & HTTPs

Georgia Tech

Path Diversity of Global Censorship Measurement

Georgia Tech

Path Diversity of Global Censorship Measurement

- Observed two modes

Georgia Tech

Path Diversity of Global Censorship Measurement

- Observed two modes
  - **Source IP** had higher impact on path density, than source port

Georgia Tech

Path Diversity of Global Censorship Measurement

- Observed two modes
  - **Source IP** had higher impact on path density, than source port
  - Both **Source IP and Source Port** had similar impact

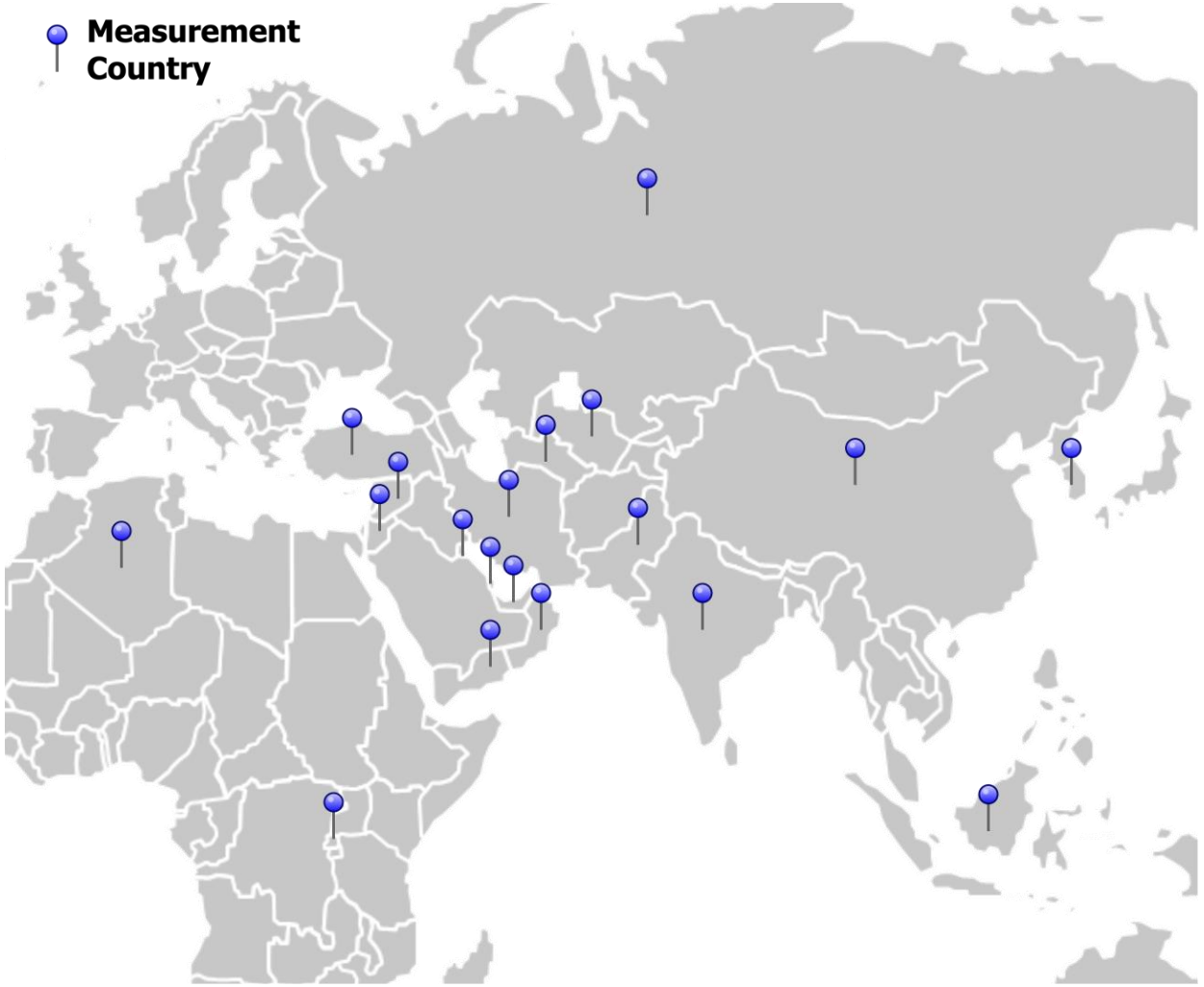
Path Diversity of Global Censorship Measurement

- Observed two modes
  - **Source IP** had higher impact on path density, than source port
  - Both **Source IP and Source Port** had similar impact

- Also observed **differing impact across protocols** for some countries

High-Level impact of ECMP on global censorship measurement results?

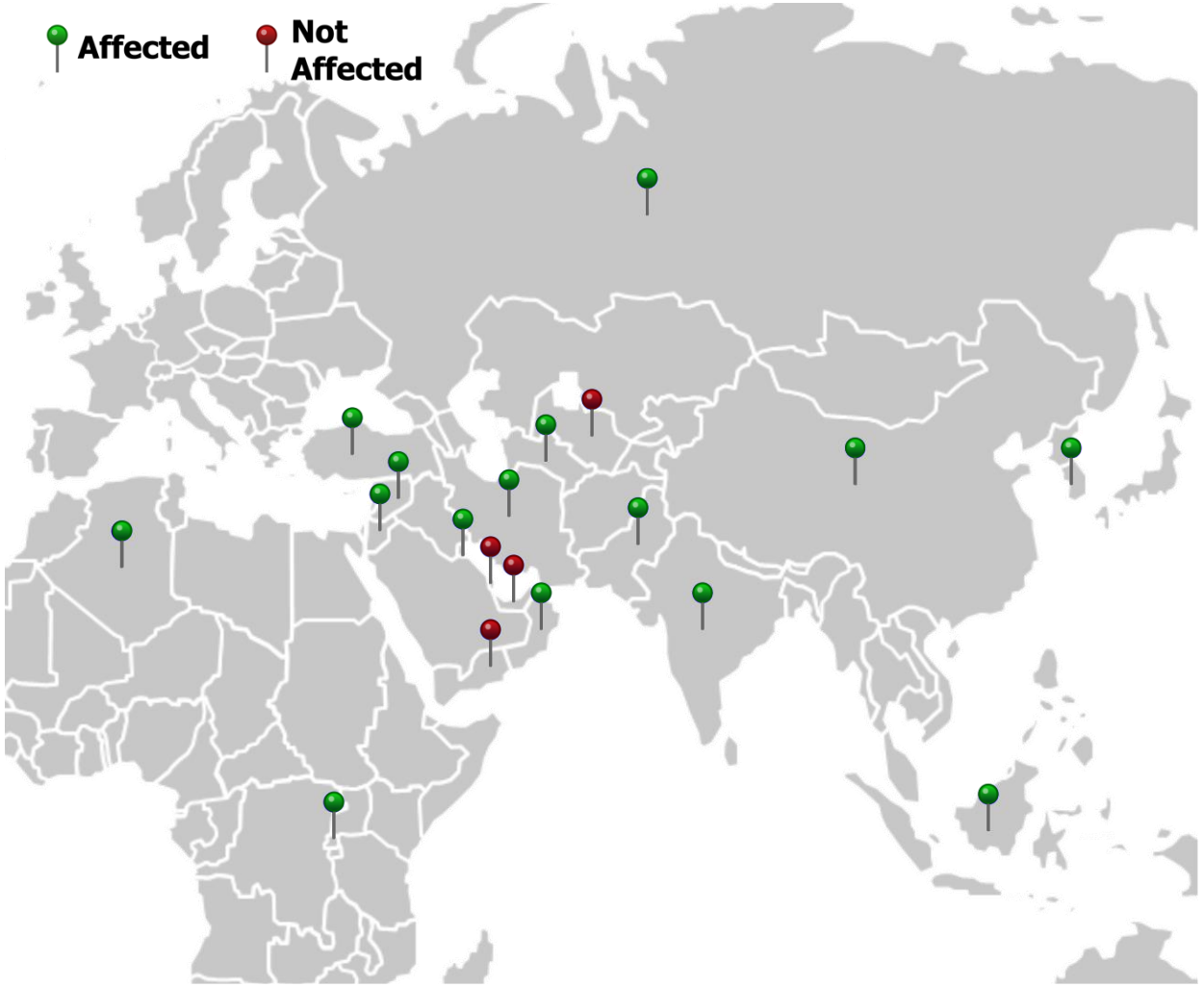
- **17 out 21** countries impacted



Georgia Tech

High-Level impact of ECMP on global censorship measurement results?

- **17 out 21** countries impacted



Georgia Tech

High-Level impact of ECMP on global censorship measurement results?

- **17 out 21** countries impacted

- Extent of impact **varies by *country* and *protocol***
  - ranges from 100% to <1% (of destinations)

Country	Protocol	Destinations	Destinations Affected (%)
India	HTTP	3289	99%
	HTTPS	3316	100%
Jordan	HTTPS	231	<1%

High-Level impact of ECMP on global censorship measurement results?

- **17 out 21** countries impacted

- Extent of impact **varies by *country* and *protocol***
  - ranges from 100% to <1% (of destinations)

- Impact can vary for **different protocols *within* a country**

Country	Protocol	Destinations	Destinations Affected (%)
Russia	DNS	105	65%
	HTTP	1808	47%
	HTTPS	2526	15%

Georgia Tech

How do changes in the packet influence censorship results?

Georgia Tech

How do changes in the packet influence censorship results?

- All source parameters had impact on censorship changes

Georgia Tech

How do changes in the packet influence censorship results?

- All source parameters had impact on censorship changes
  - Source IPs

	Context	Parameter	Impact
Source IP	China (HTTPS)	**1.2.3.8:**XXXX	
		1.2.3.23:XXXX	

Georgia Tech

How do changes in the packet influence censorship results?

- All source parameters had impact on censorship changes
  - Source IPs
  - Source IP and Source Port

	Context	Parameter	Impact
Source IP	China (HTTPS)	**1.2.3.8:**XXXX	
		1.2.3.23:XXXX	
Source IP and Source Port	Bangladesh (HTTP)	**1.2.3.8:12345**	
		1.2.3.23:56789	

Georgia Tech

How do changes in the packet influence censorship results?

- All source parameters had impact on censorship changes
  - Source IPs
  - Source IP and Source Port
  - Source Port

	Context	Parameter	Impact
Source IP	China (HTTPS)	**1.2.3.8:**XXXX	
		1.2.3.23:XXXX	
Source IP and Source Port	Bangladesh (HTTP)	**1.2.3.8:12345**	
		1.2.3.23:56789	
Source Port	Iran (DNS)	1.2.3.X:**12345**	
		1.2.3.X:**56789**	

Georgia Tech

How do changes in the packet influence censorship results?

- All source parameters had impact on censorship changes
  - Source IPs
  - Source IP and Source Port
  - Source Port
- Indicates different forms of ECMP routing

	Context	Parameter	Impact
Source IP	China (HTTPS)	**1.2.3.8:**XXXX	
		1.2.3.23:XXXX	
Source IP and Source Port	Bangladesh (HTTP)	**1.2.3.8:12345**	
		1.2.3.23:56789	
Source Port	Iran (DNS)	1.2.3.X:**12345**	
		1.2.3.X:**56789**	

Georgia Tech

What underlying network structures cause change?

Georgia Tech

What underlying network structures cause change?

- We find 3 effects contributing to change

Georgia Tech

What underlying network structures cause change?

- We find 3 effects contributing to change

1. **ECMP Routing (Inter-AS or Intra-AS) exercising possibly failed/misconfigured censoring nodes**

Georgia Tech

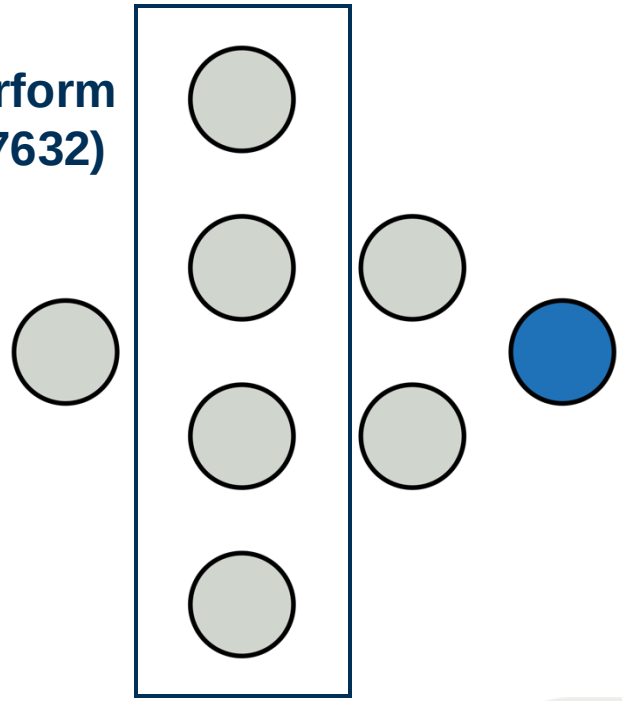
What underlying network structures cause change?

- We find 3 effects contributing to change

1. **ECMP Routing (Inter-AS or Intra-AS) exercising possibly failed/misconfigured censoring nodes**

Indonesia (HTTPS)

AS known to perform censorship (AS7632)

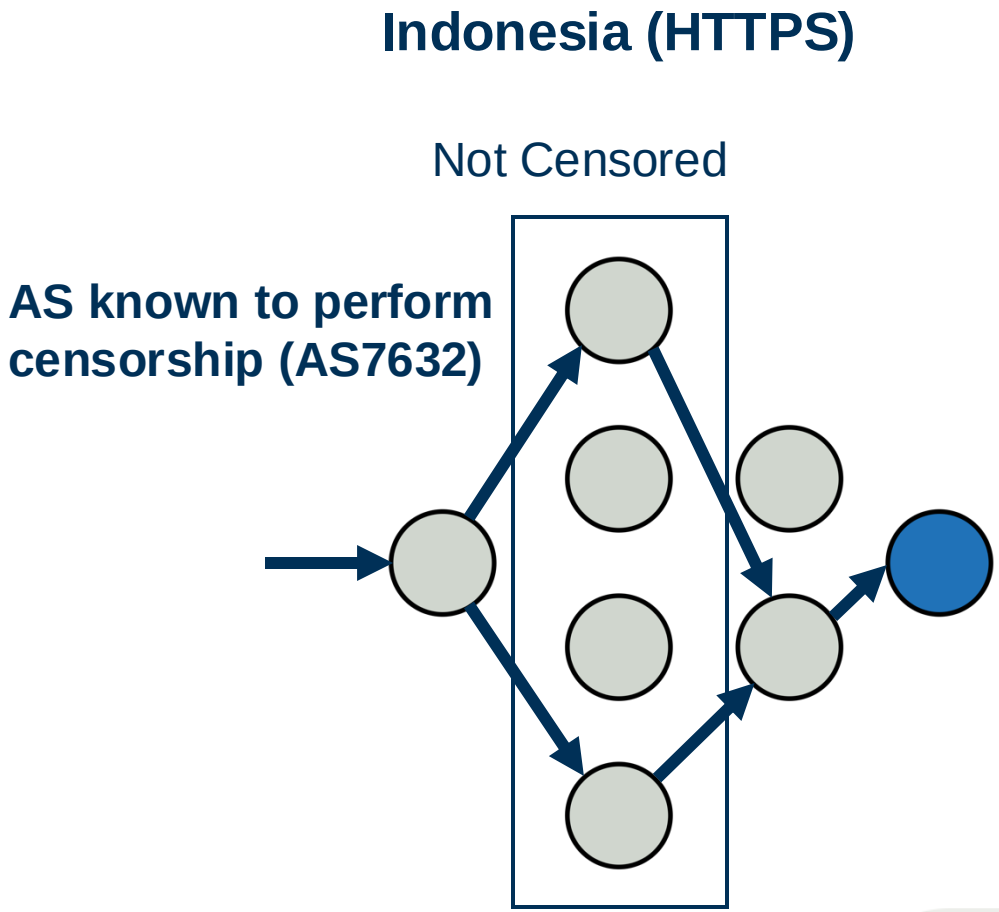


Georgia Tech

What underlying network structures cause change?

- We find 3 effects contributing to change

1. **ECMP Routing (Inter-AS or Intra-AS) exercising possibly failed/misconfigured censoring nodes**

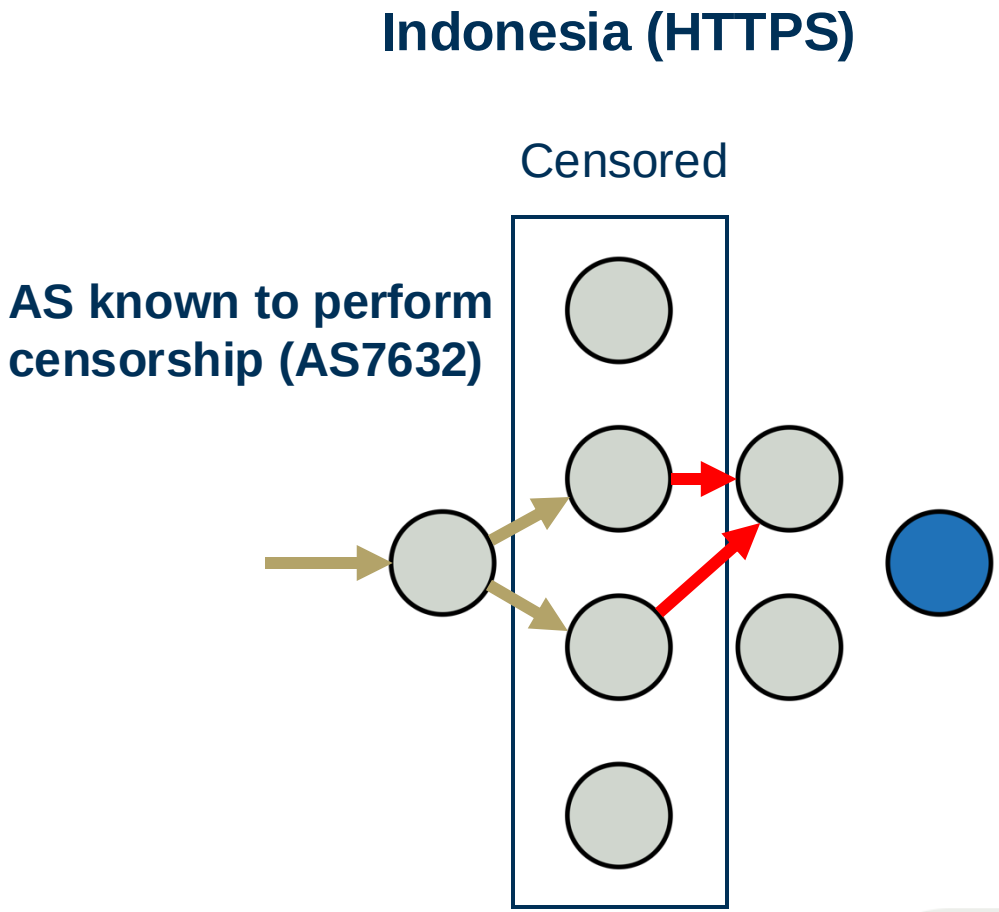


Georgia Tech

What underlying network structures cause change?

- We find 3 effects contributing to change

1. **ECMP Routing (Inter-AS or Intra-AS) exercising possibly failed/misconfigured censoring nodes**



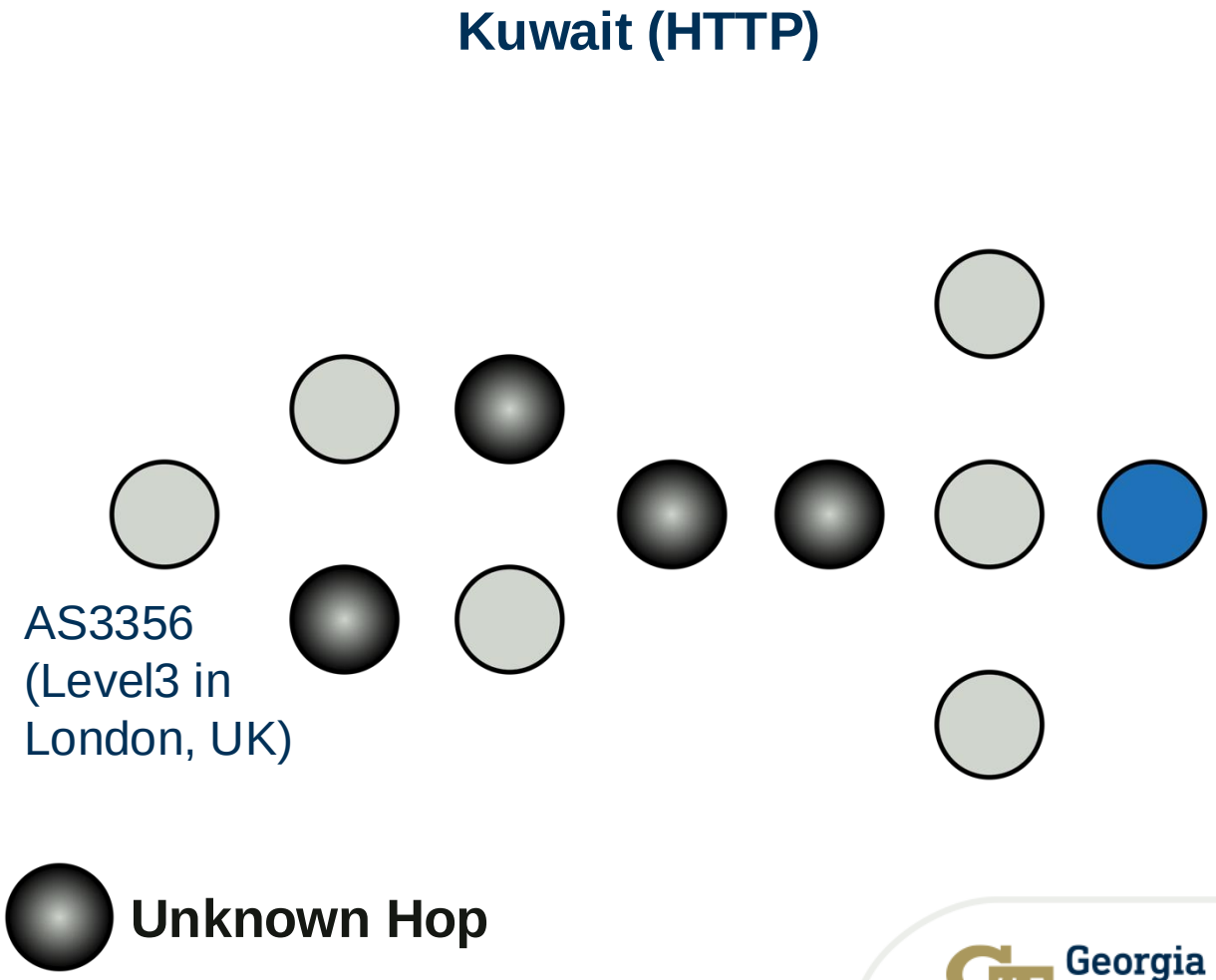
Georgia Tech

What underlying network structures cause change?

- We find 3 effects contributing to change

1. ECMP Routing (Inter-AS or Intra-AS) exercising possibly failed/misconfigured censoring nodes

2. **ECMP routing around censorship**

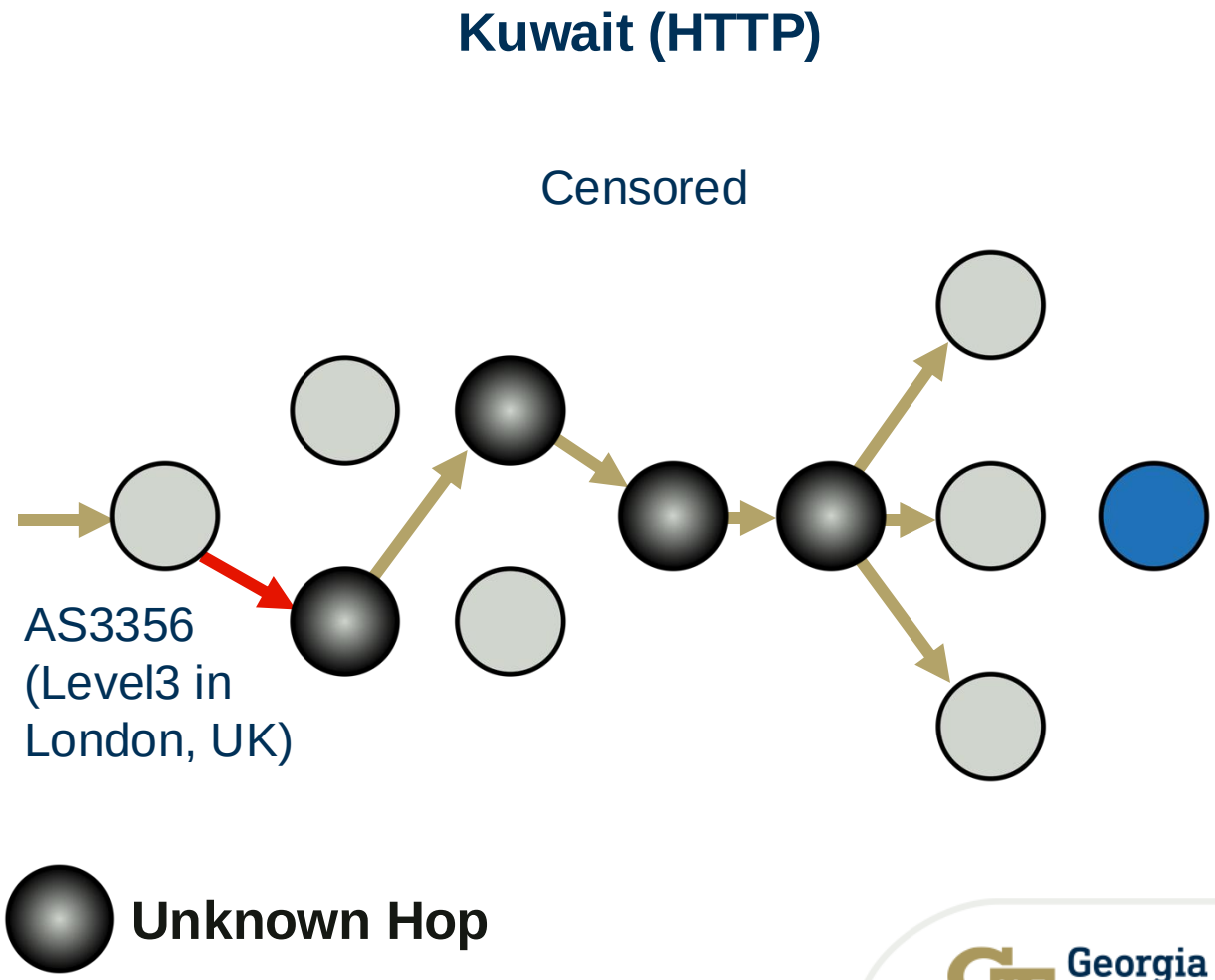


What underlying network structures cause change?

- We find 3 effects contributing to change

1. ECMP Routing (Inter-AS or Intra-AS) exercising possibly failed/misconfigured censoring nodes

2. **ECMP routing around censorship**

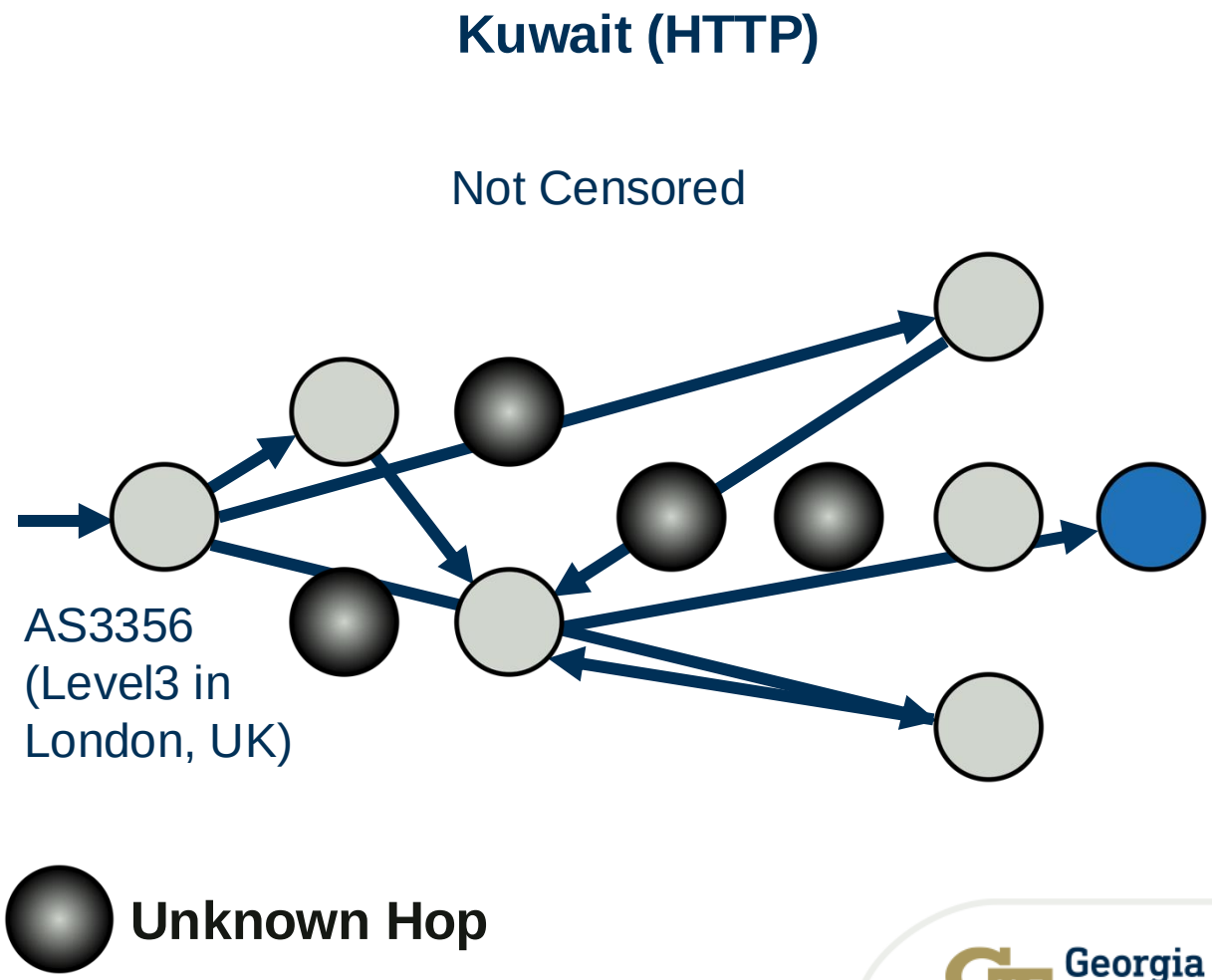


What underlying network structures cause change?

- We find 3 effects contributing to change

1. ECMP Routing (Inter-AS or Intra-AS) exercising possibly failed/misconfigured censoring nodes

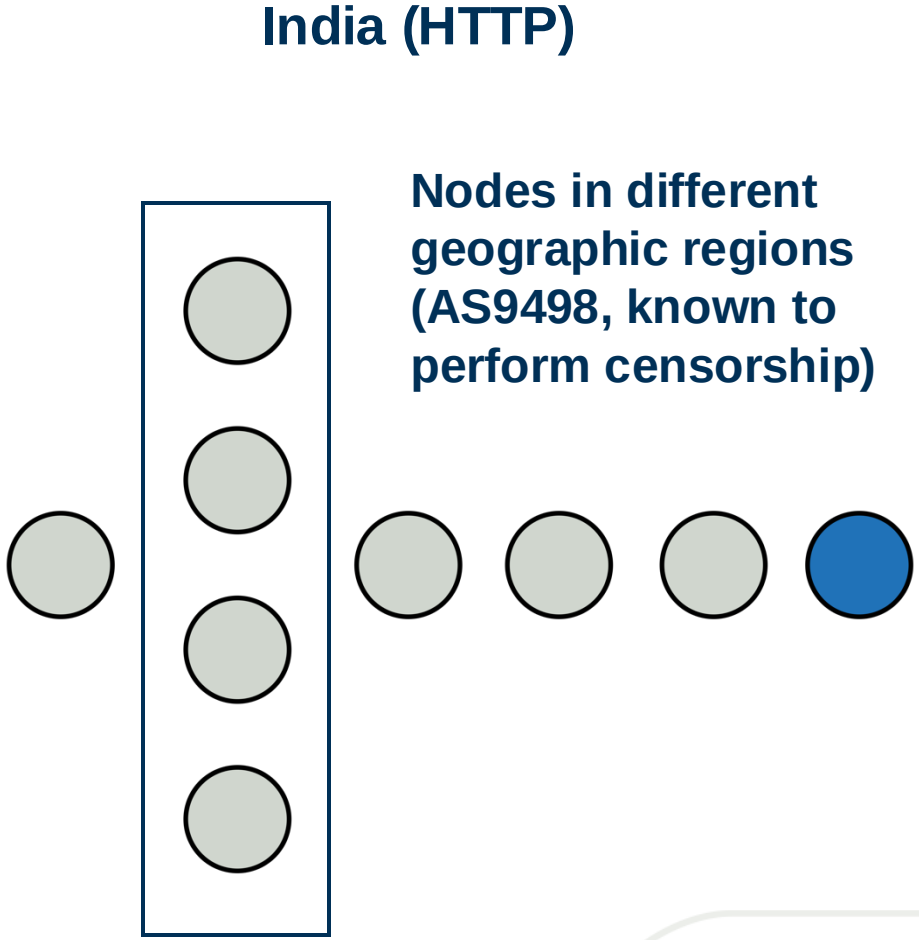
2. **ECMP routing around censorship**



What underlying network structures cause change?

- We find 3 effects contributing to change

1. ECMP Routing (Inter-AS or Intra-AS) exercising possibly failed/misconfigured censoring nodes
2. ECMP routing around censorship
3. **ECMP routing through geographically different regions with different censorship behavior**

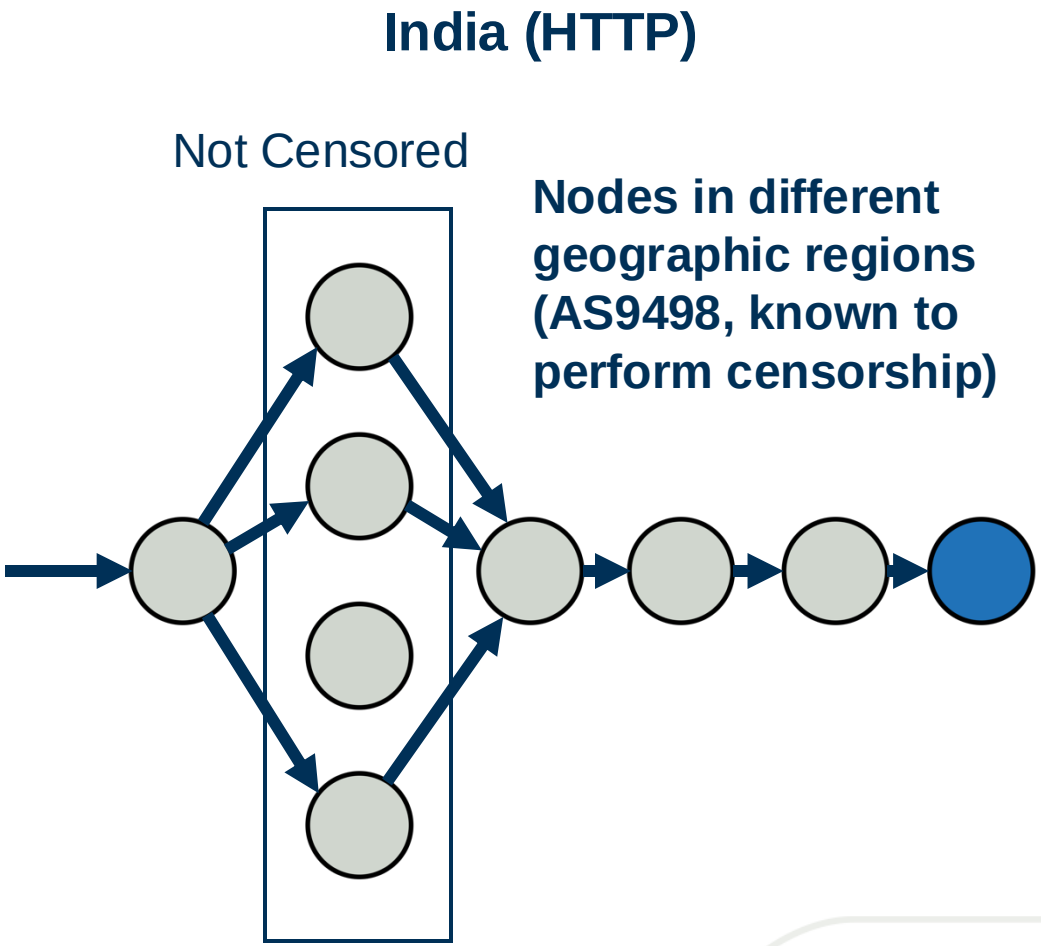


Georgia Tech

What underlying network structures cause change?

- We find 3 effects contributing to change

1. ECMP Routing (Inter-AS or Intra-AS) exercising possibly failed/misconfigured censoring nodes
2. ECMP routing around censorship
3. **ECMP routing through geographically different regions with different censorship behavior**

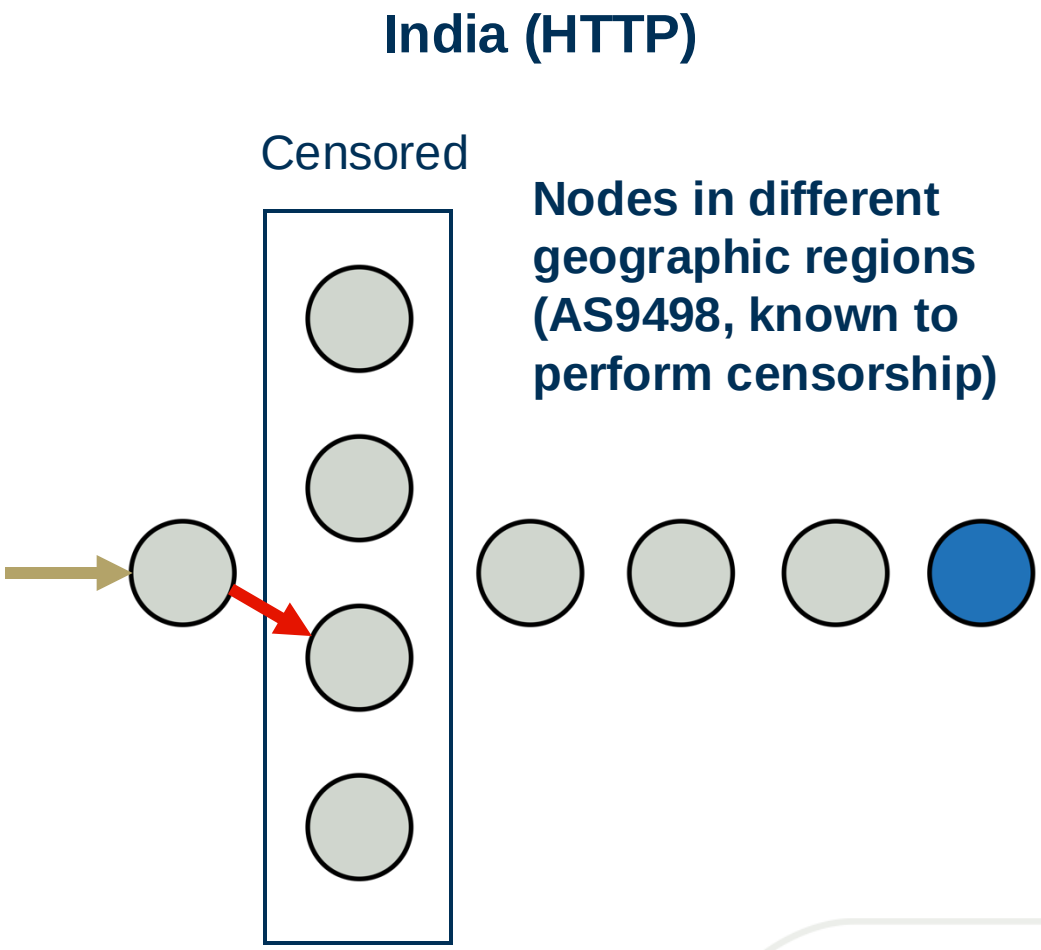


Georgia Tech

What underlying network structures cause change?

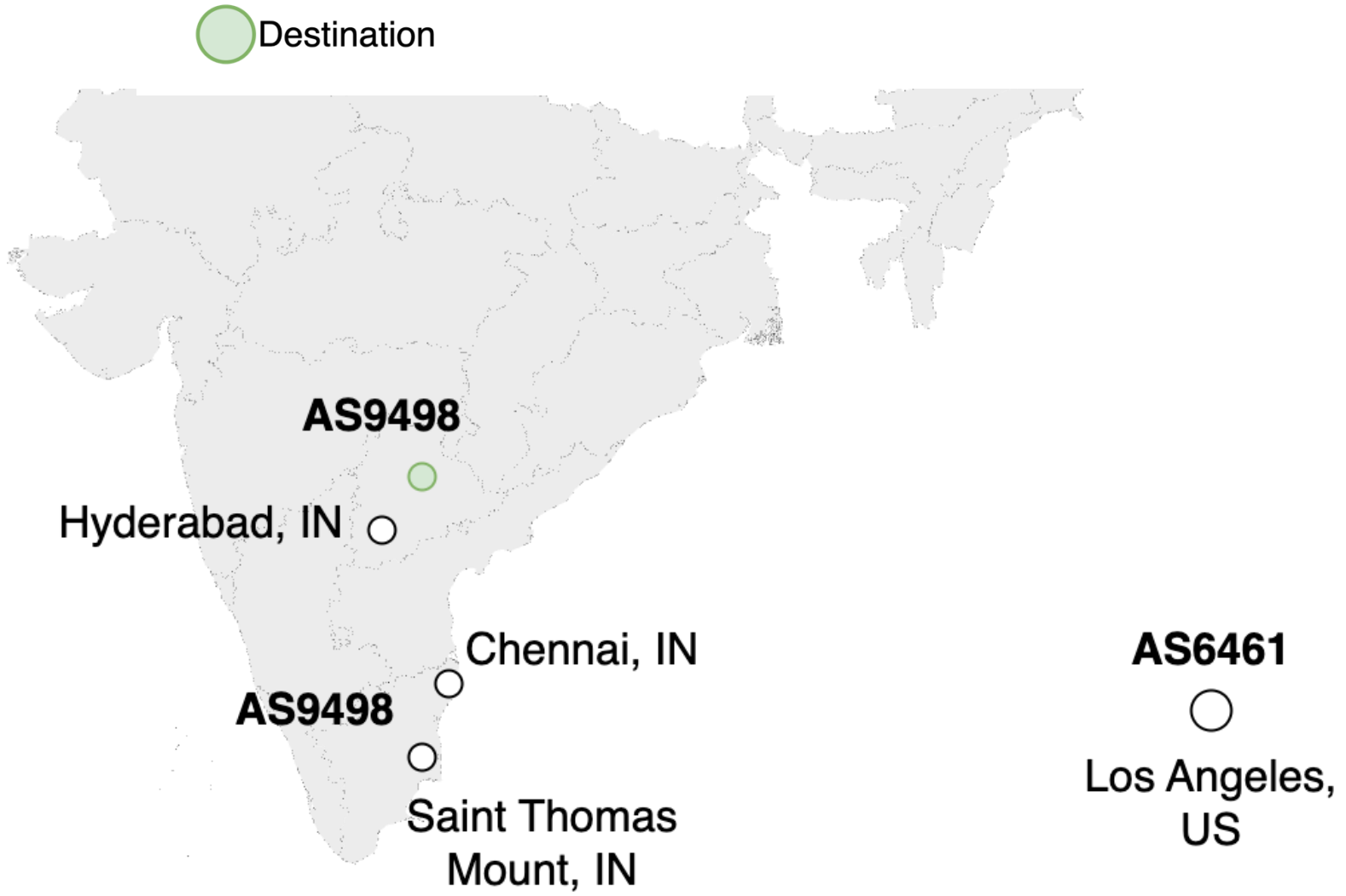
- We find 3 effects contributing to change

1. ECMP Routing (Inter-AS or Intra-AS) exercising possibly failed/misconfigured censoring nodes
2. ECMP routing around censorship
3. **ECMP routing through geographically different regions with different censorship behavior**



Georgia Tech

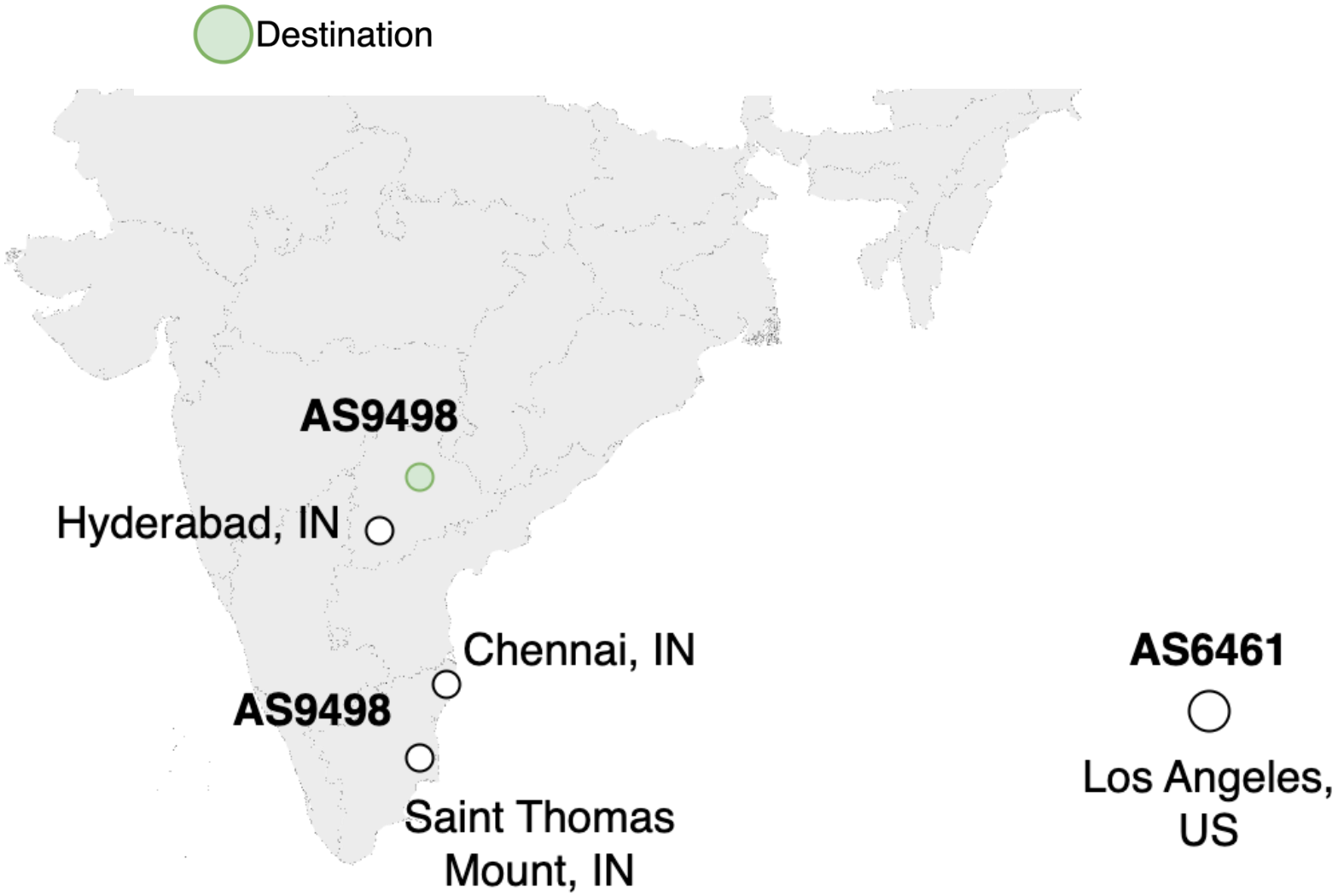
ECMP routing through geographically different regions with different censorship behavior



Georgia Tech

ECMP routing through geographically different regions with different censorship behavior

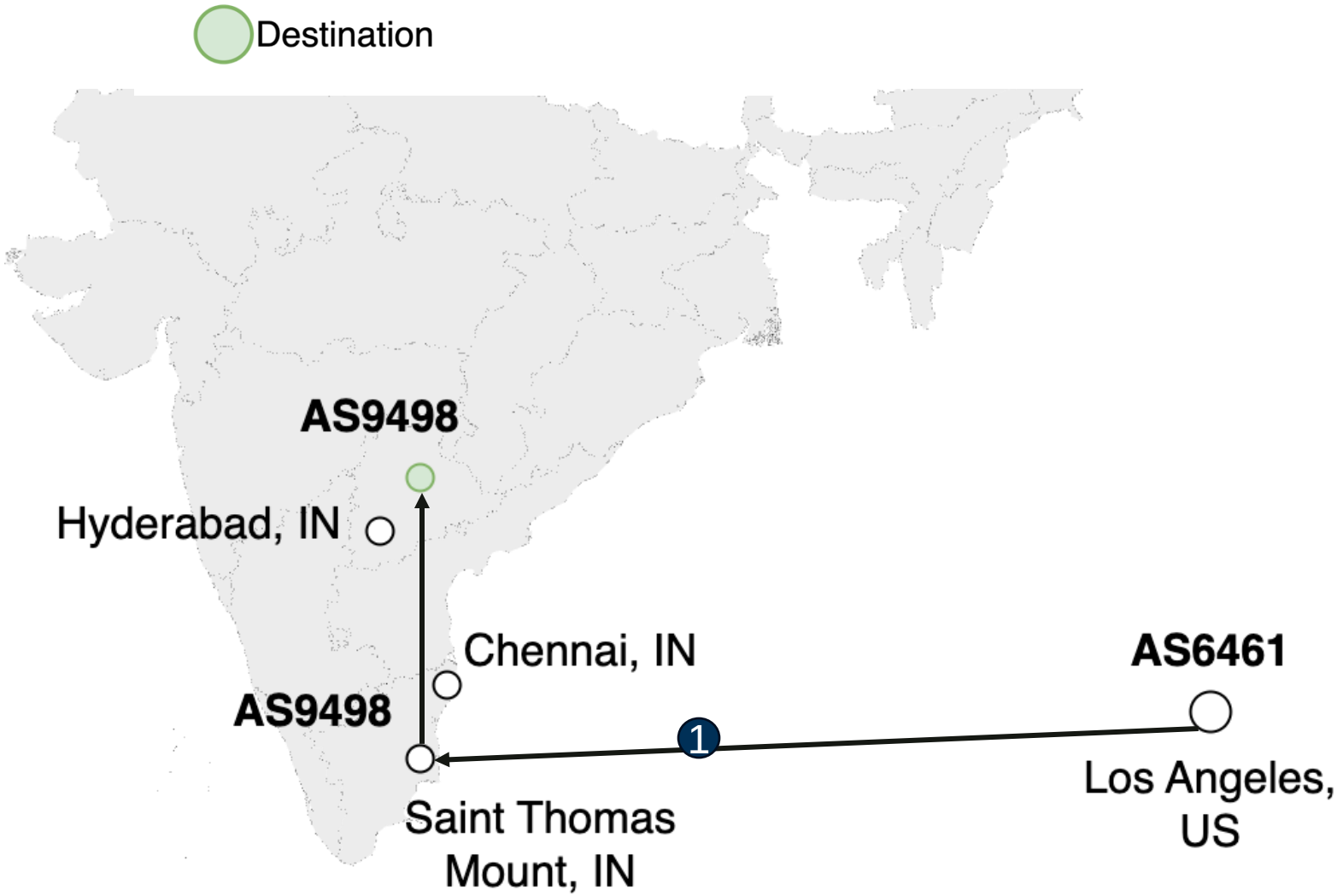
- Sensitive packet takes ***geographically different*** *paths* into the country



Georgia Tech

ECMP routing through geographically different regions with different censorship behavior

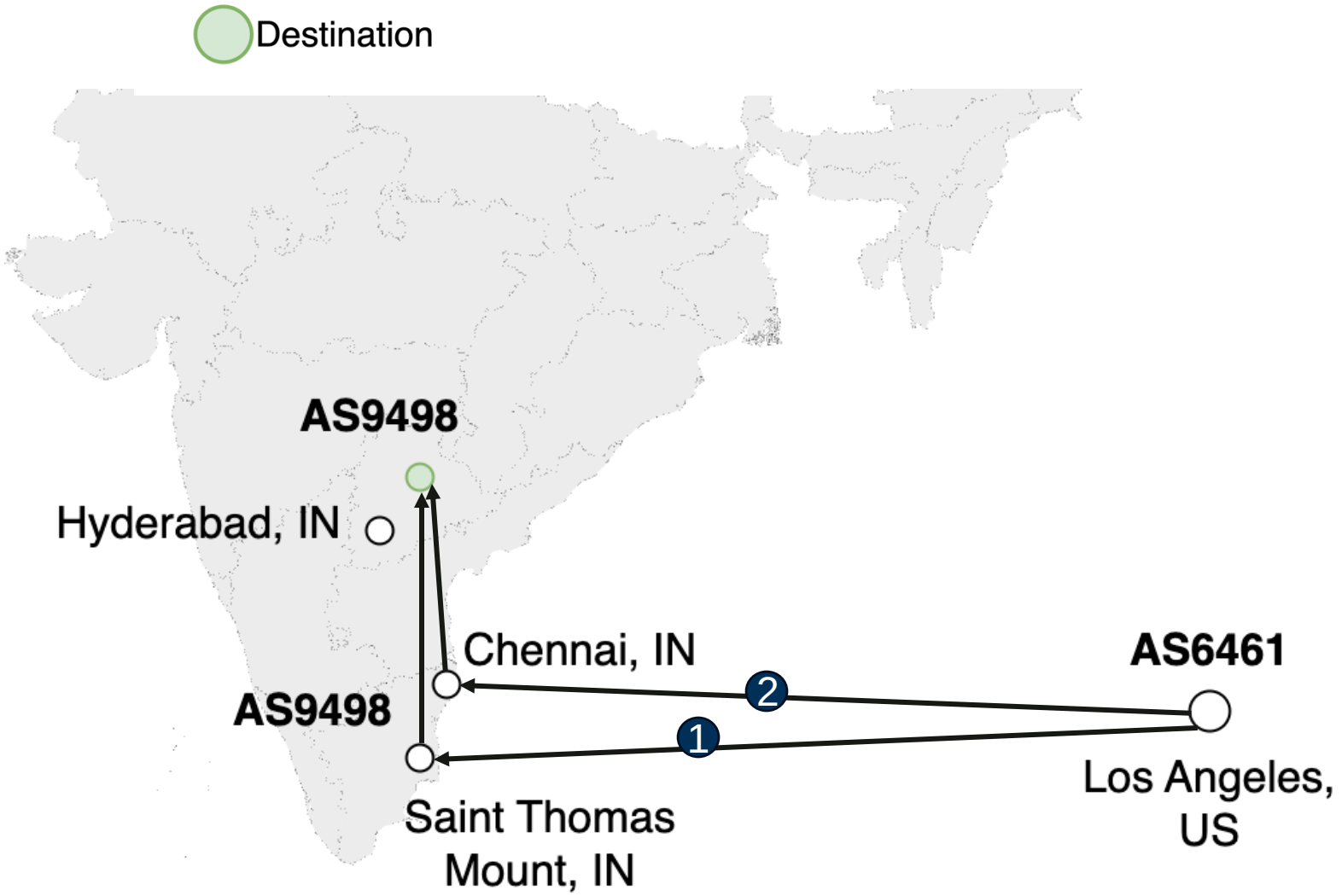
- Sensitive packet takes ***geographically different*** *paths* into the country



Georgia Tech

ECMP routing through geographically different regions with different censorship behavior

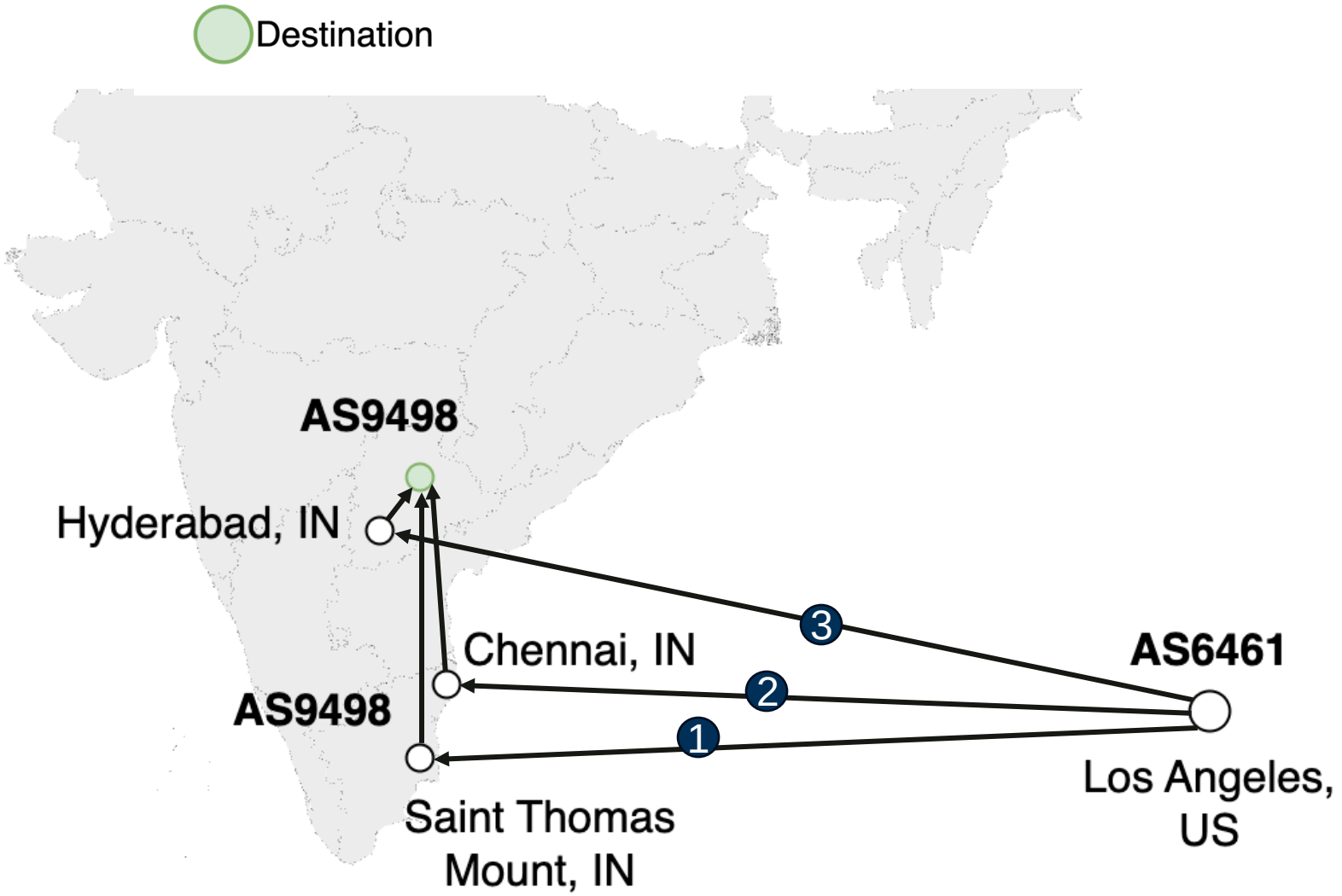
- Sensitive packet takes ***geographically different*** *paths* into the country



Georgia Tech

ECMP routing through geographically different regions with different censorship behavior

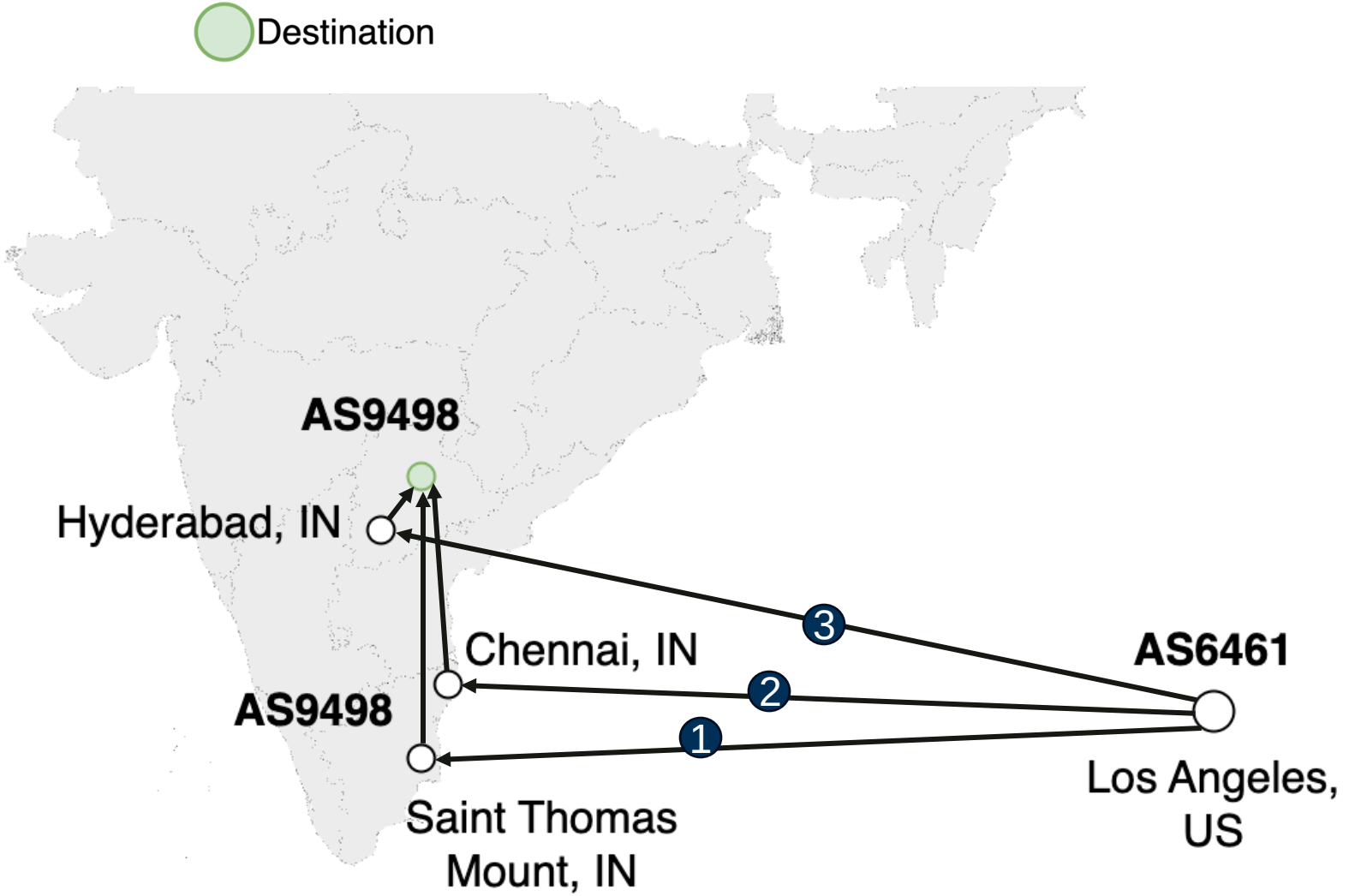
- Sensitive packet takes ***geographically different*** *paths* into the country



Georgia Tech

ECMP routing through geographically different regions with different censorship behavior

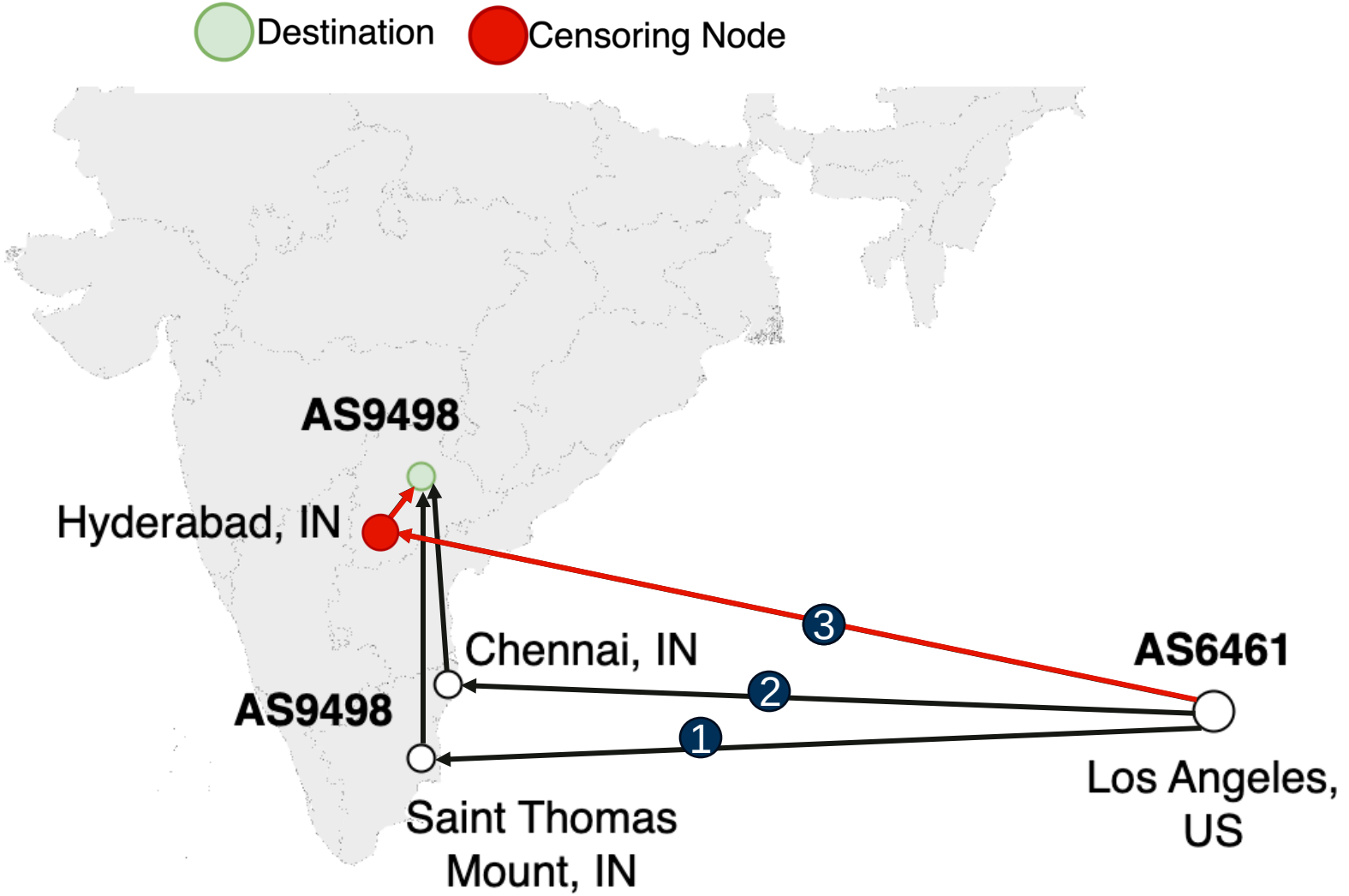
- Sensitive packet takes ***geographically different*** *paths* into the country
- Some of the paths experience censorship, while others do not



Georgia Tech

ECMP routing through geographically different regions with different censorship behavior

- Sensitive packet takes ***geographically different*** *paths* into the country
- Some of the paths experience censorship, while others do not



Georgia Tech

Discussion and Recommendation

Georgia Tech

Discussion and Recommendation

- Studies must account for censorship ***along the path*** rather than ***at the destination***

Georgia Tech

Discussion and Recommendation

- Studies must account for censorship ***along the path*** rather than ***at the destination***

- Measurement over **HTTP, HTTPS and DNS** must account for ECMP routing

Georgia Tech

Discussion and Recommendation

- Studies must account for censorship ***along the path*** rather than ***at the destination***
- Measurement over **HTTP, HTTPS and DNS** must account for ECMP routing
- Diversity in **source IP** and **source port** is critical

Discussion and Recommendation

- Studies must account for censorship ***along the path*** rather than ***at the destination***
- Measurement over **HTTP, HTTPS and DNS** must account for ECMP routing
- Diversity in **source IP** and **source port** is critical
- **Not all non-determinism** is due to ECMP routing, but must be taken into account

Discussion and Recommendation

- Studies must account for censorship ***along the path*** rather than ***at the destination***
- Measurement over **HTTP, HTTPS and DNS** must account for ECMP routing
- Diversity in **source IP** and **source port** is critical
- **Not all non-determinism** is due to ECMP routing, but must be taken into account

Thank you!
Questions?

: abhaskar@gatech.edu
: abhaskarv

Georgia Tech