Censored Planet:

Measuring Internet Censorship Globally and Continuously

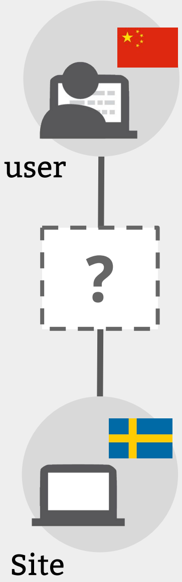
Roya Ensafi

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:

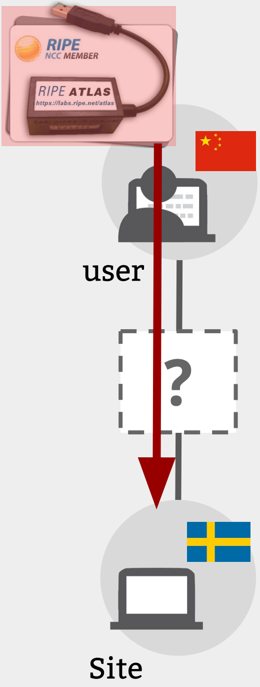
- How can we detect whether pairs of hosts around the world can talk to each other?

STATE OF THE ART:

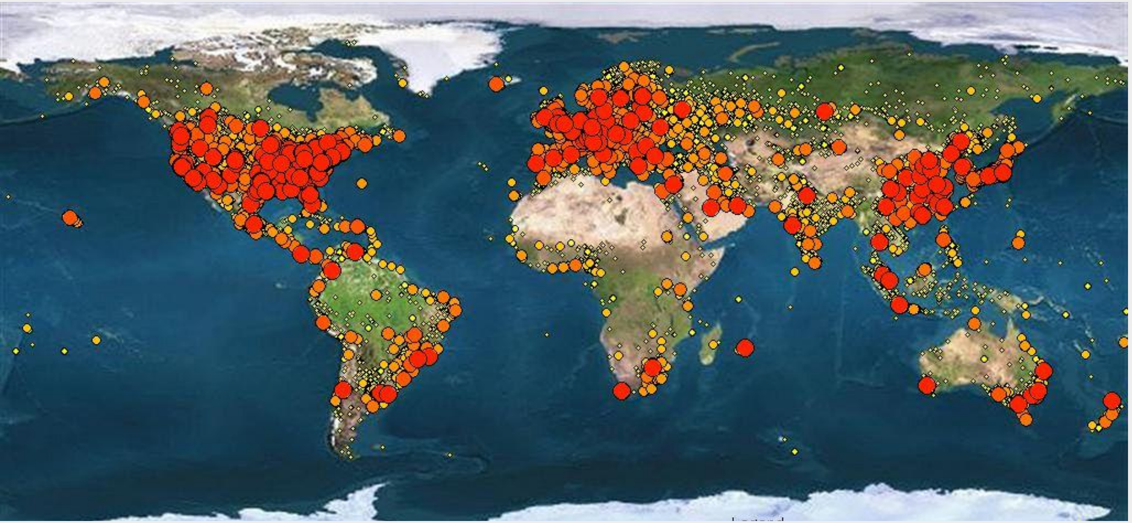
- Deploy hardware or software at hosts (RIPE Atlas, OONI probe)
- Ask people on the ground, or use VPNs, or research networks (PlanetLab)

THREE KEY CHALLENGES:

Coverage, ethics, and continuity



Thinking Like an “Attacker”…



140 million public live IPv4 addresses

These machines blindly follow Internet protocol rules such as TCP/IP.

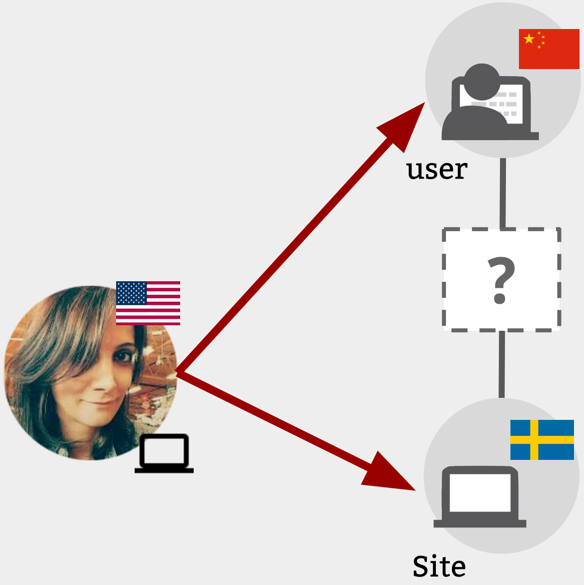
How can we leverage standard protocol behaviors to detect whether two distant hosts can communicate?

Measuring Internet Censorship Globally… Remotely!

PROBLEM:

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

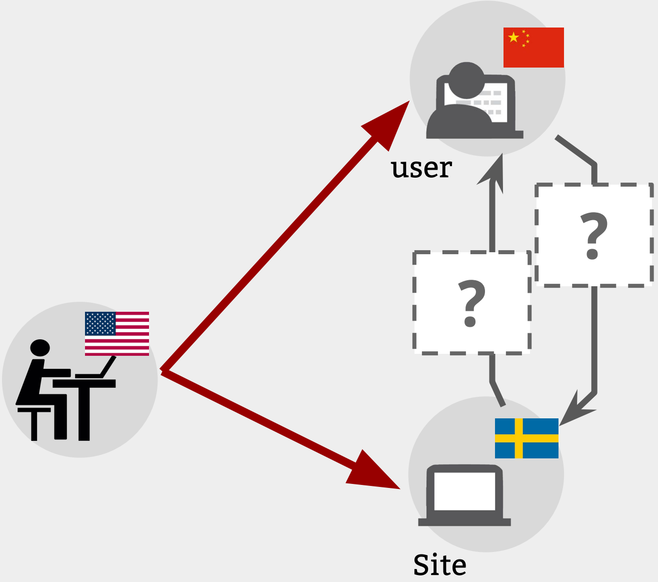
…from somewhere else in the world?



Spooky Scan

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**

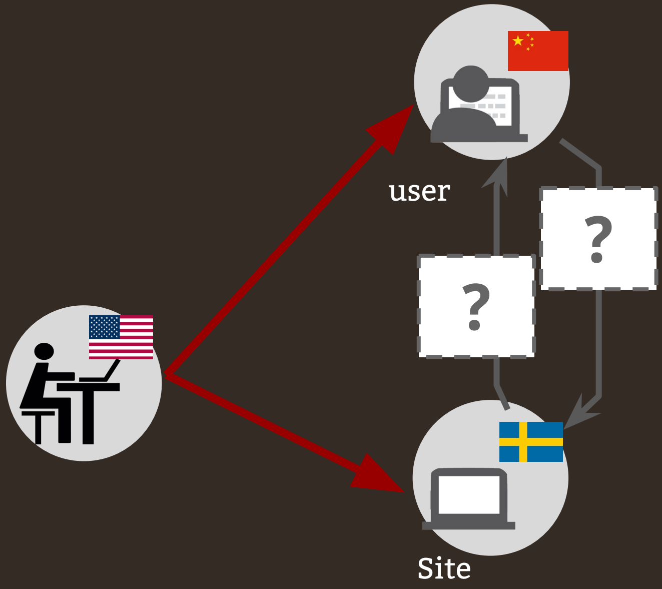


* **TCP Idle Scan** Antirez, (Bugtraq 1998)
* **Detecting Intentional Packet Drops on the Internet via TCP/IP Side Channels**
Roya Ensafi, Knockel, Alexander, and Crandall (PAM ’14)
* **Idle Port Scanning and Non-interference Analysis of Network Protocol Stacks Using Model Checking**
Roya Ensafi, Park, Kapur, and Crandall (Usenix Security 2010)

Augur

Augur is a follow up system that uses the same TCP/IP side channels to detect blocking from off-path.

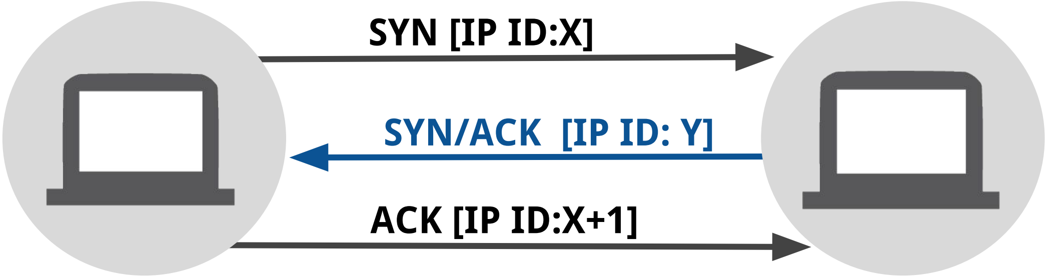
Goal: **Scalable, ethical, and statistically robust system to continuously detect blocking.**

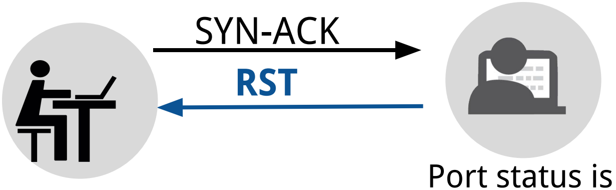


* Augur: Internet-Wide Detection of Connectivity Disruption
P. Pearce*, R. Ensafi*, F. Li, N. Feamster, V. Paxson
(* joint first authors)

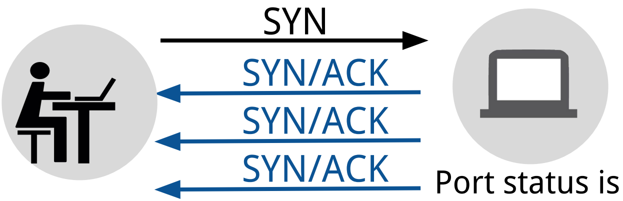
TCP/IP

TCP Handshake:





Port status is **open/closed**



Port status is **open**

Spooky Scan Requirements

“User” (Reflector)

Must maintain a global value for IP ID

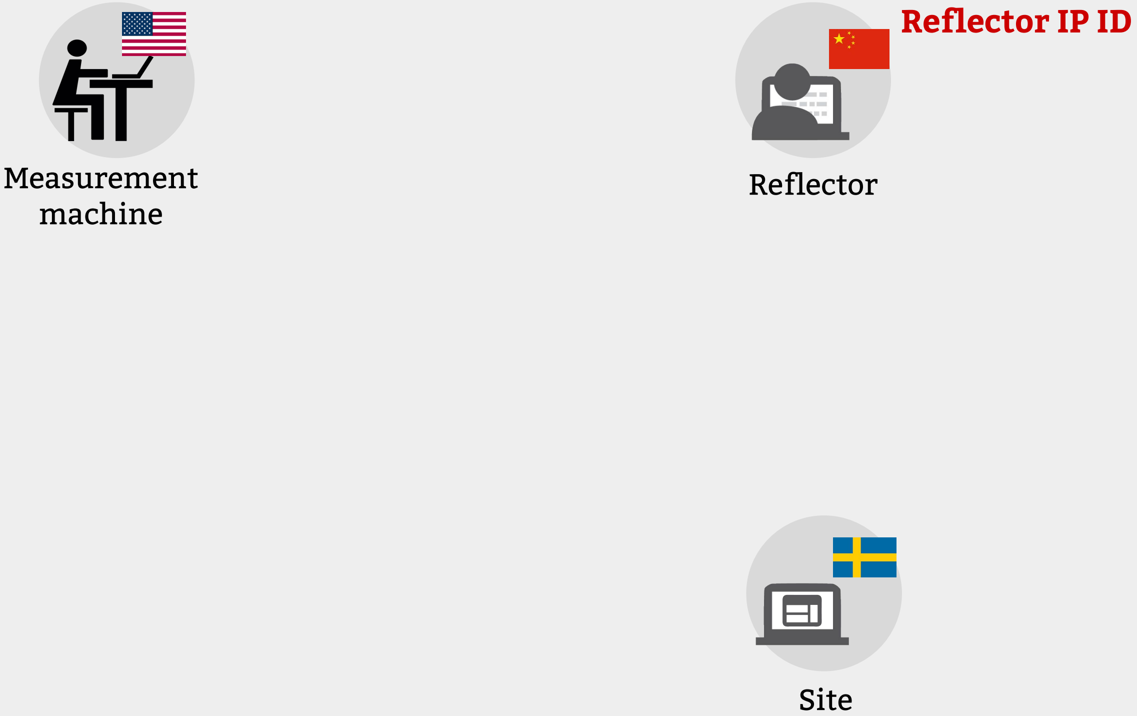
Site

Open port and retransmitting SYN-ACKs

Measurement Machine

Must be able to spoof packets

Spooky Scan



Measurement machine

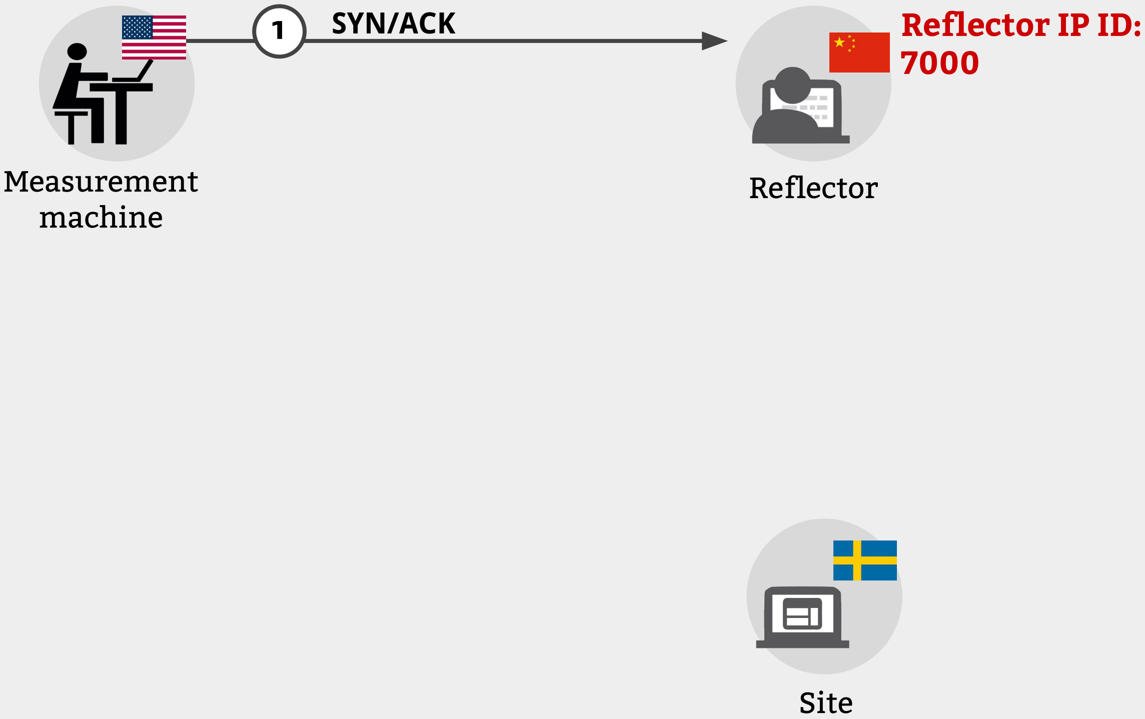
Reflector

Reflector IP ID

Site

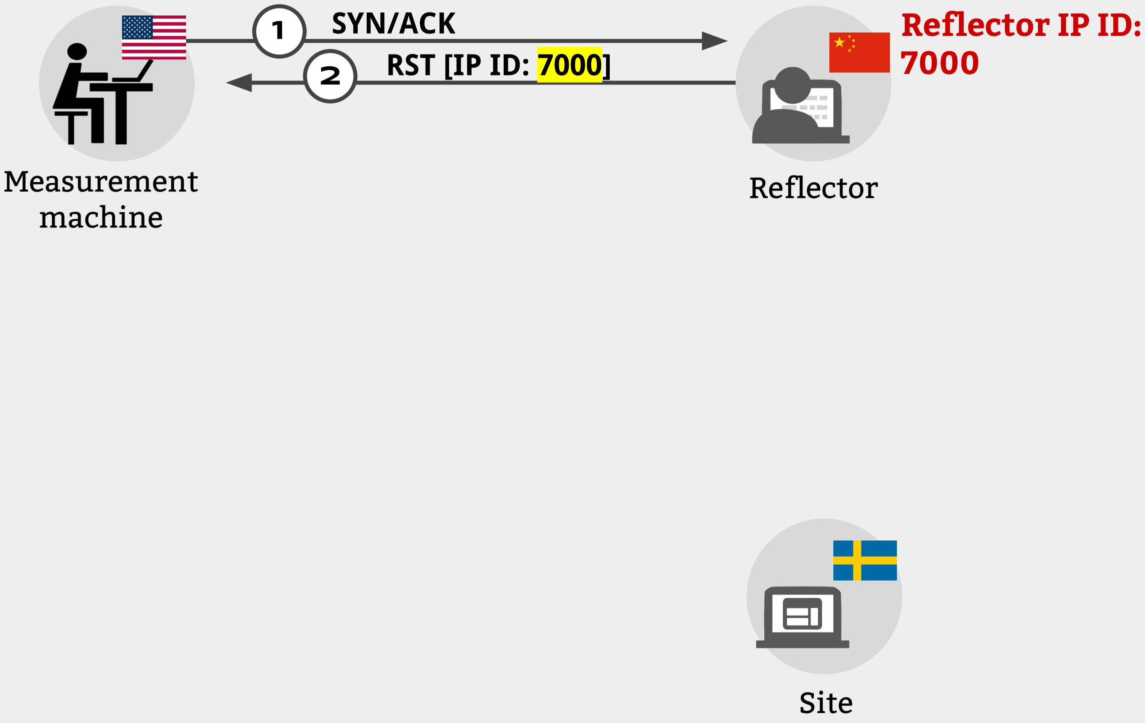
Spooky Scan

No direction blocked



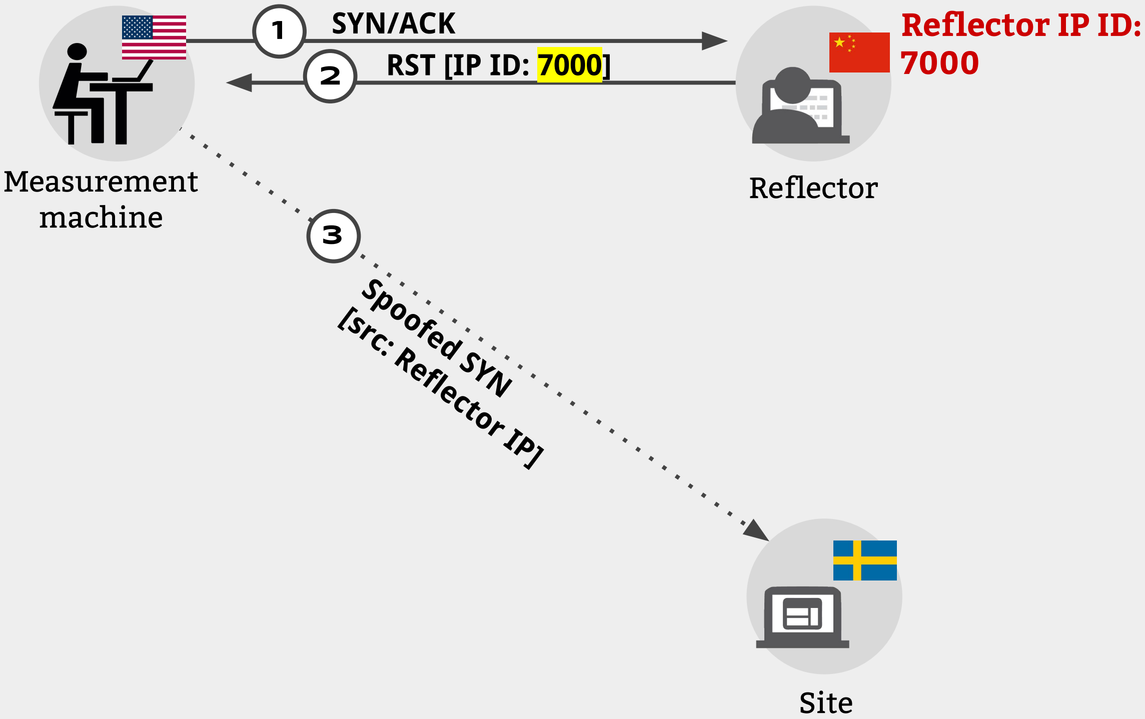
Spooky Scan

No direction blocked



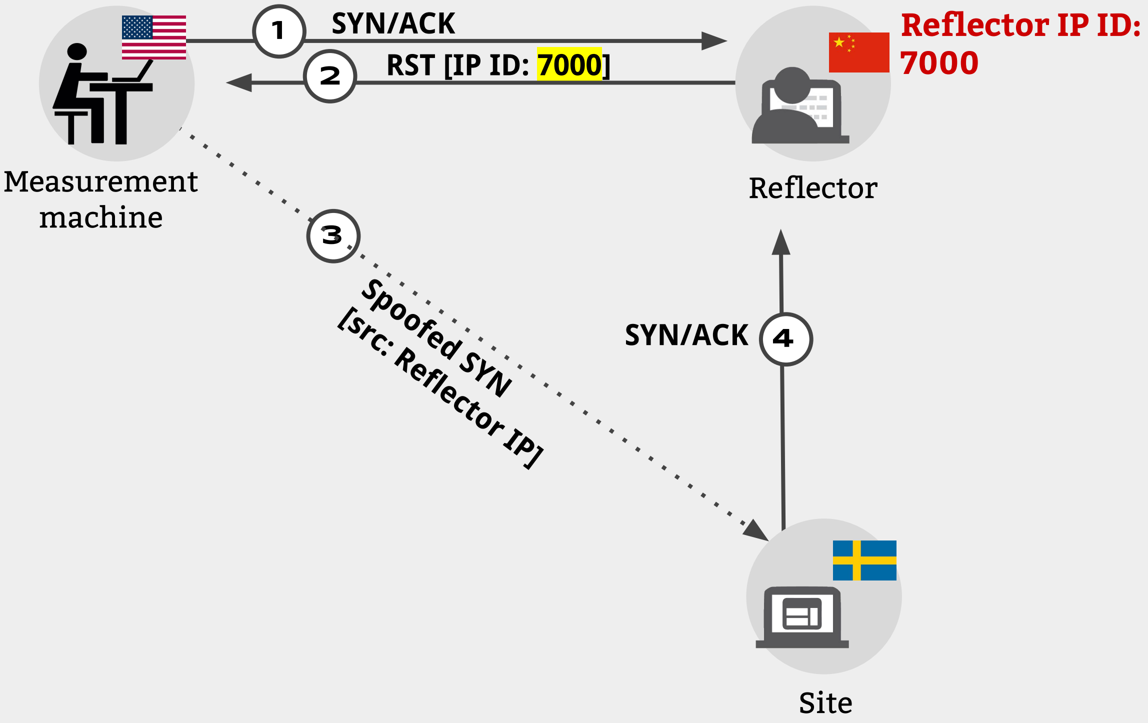
Spooky Scan

No direction blocked



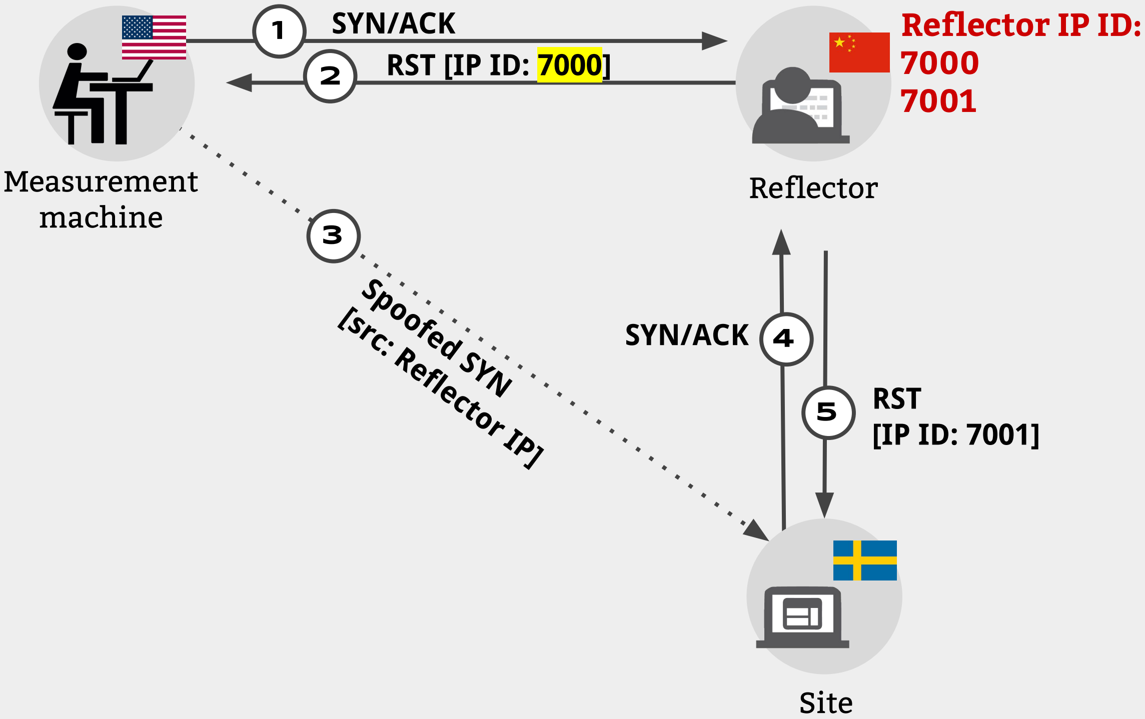
Spooky Scan

No direction blocked



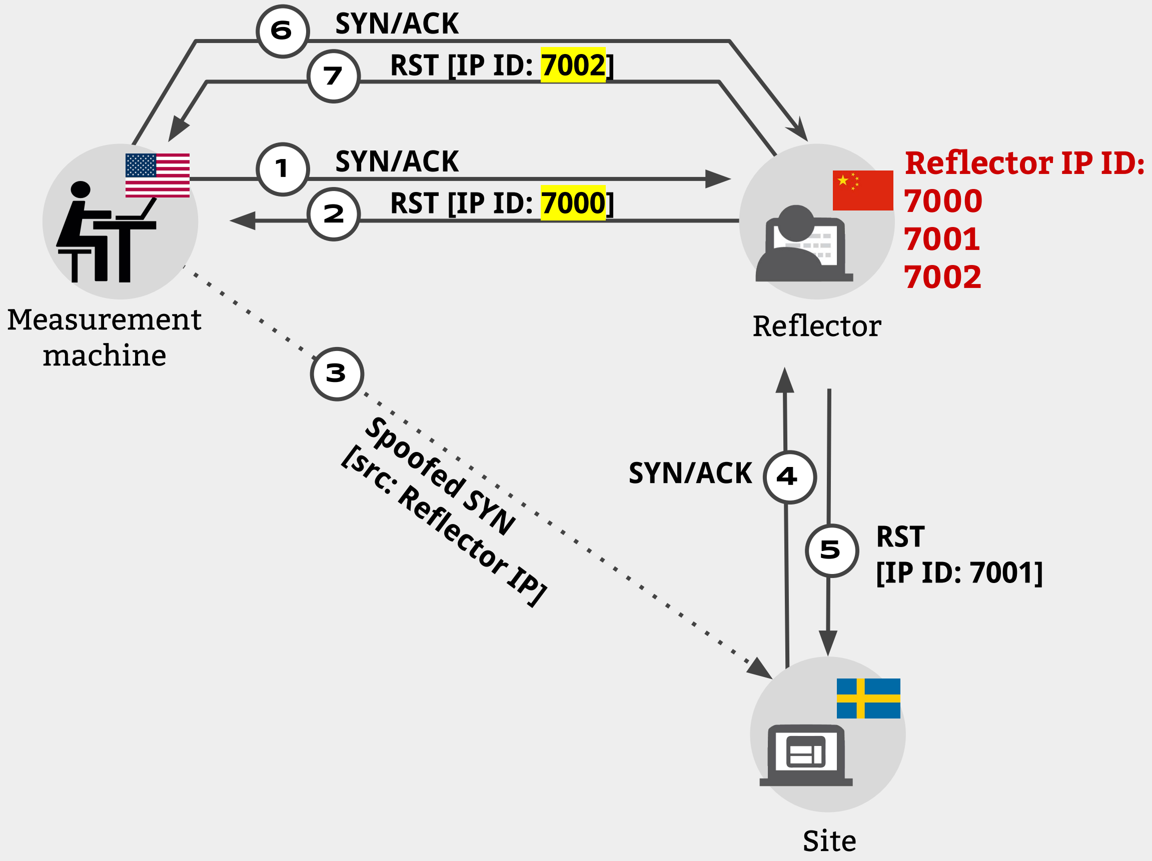
Spooky Scan

No direction blocked



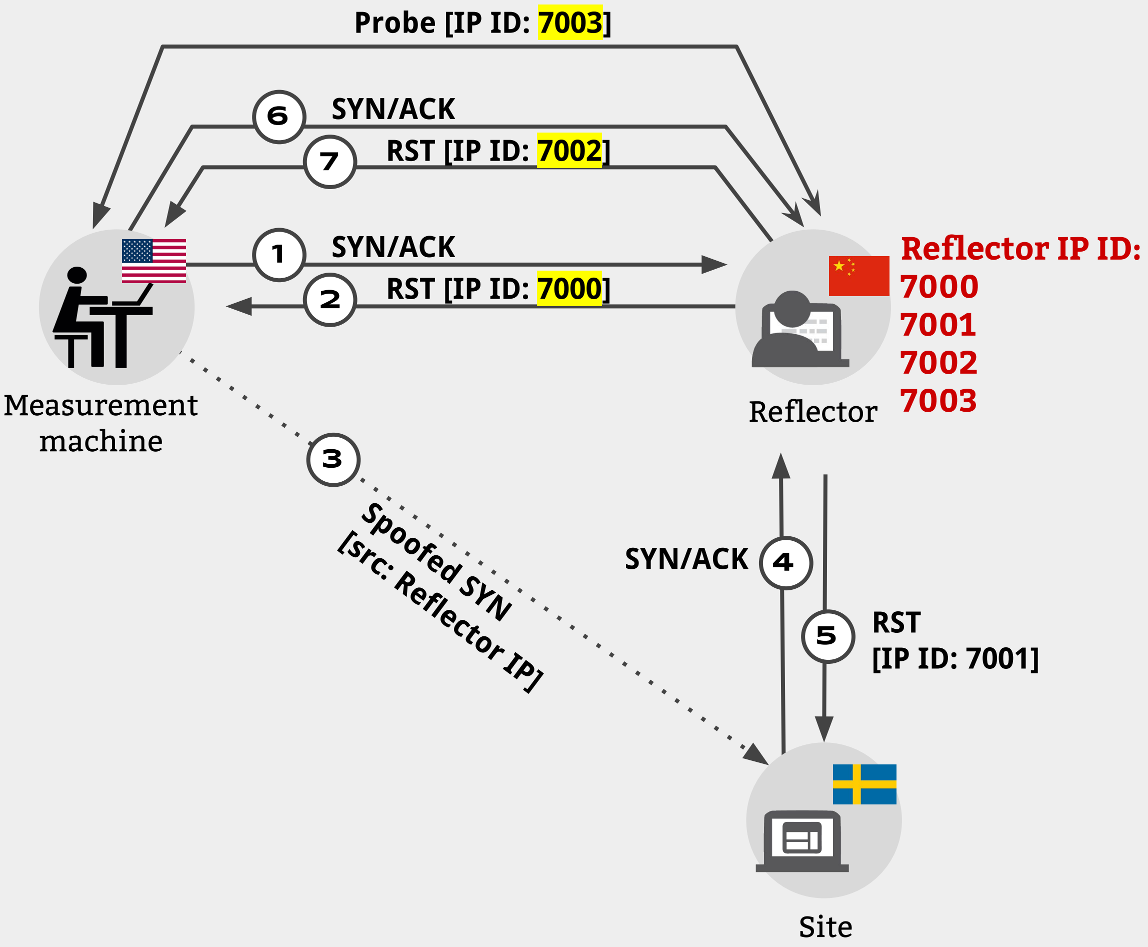
Spooky Scan

No direction blocked



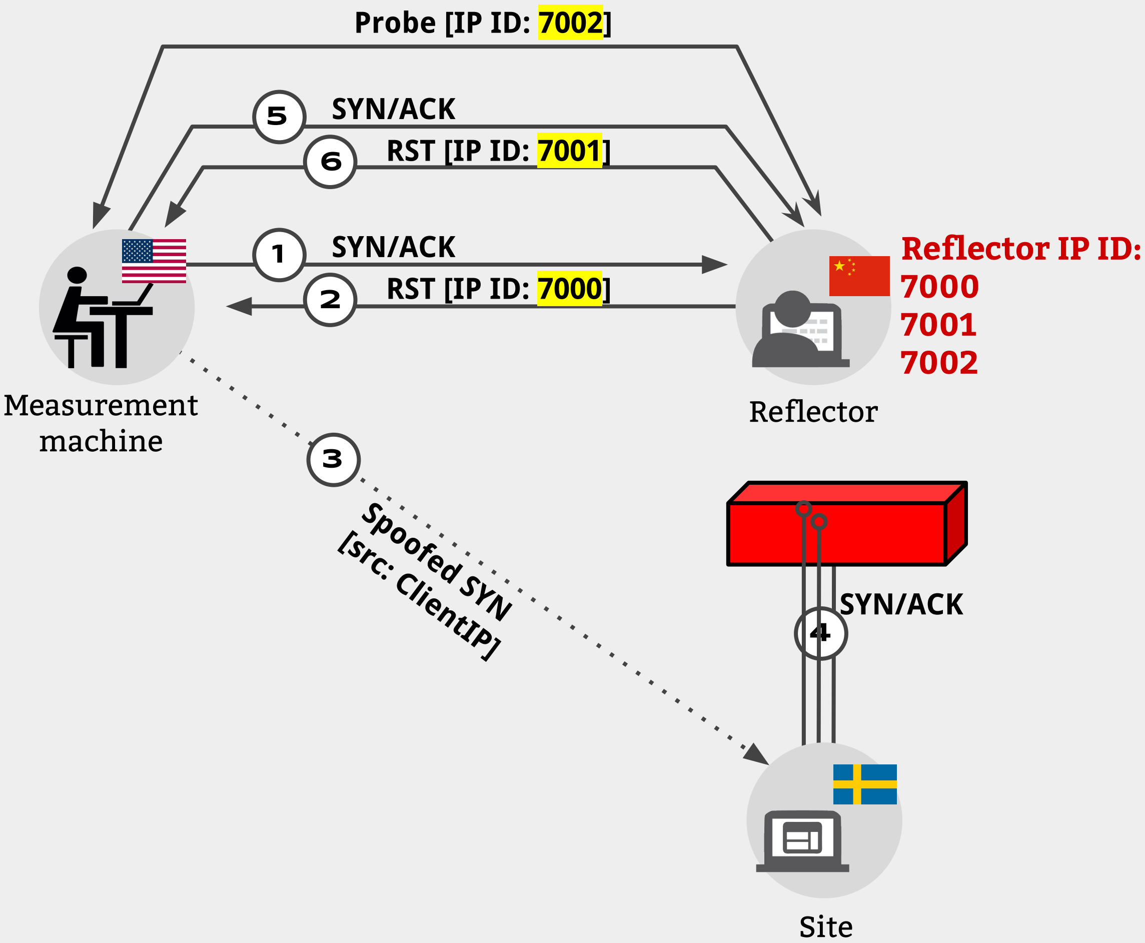
Spooky Scan

No direction blocked



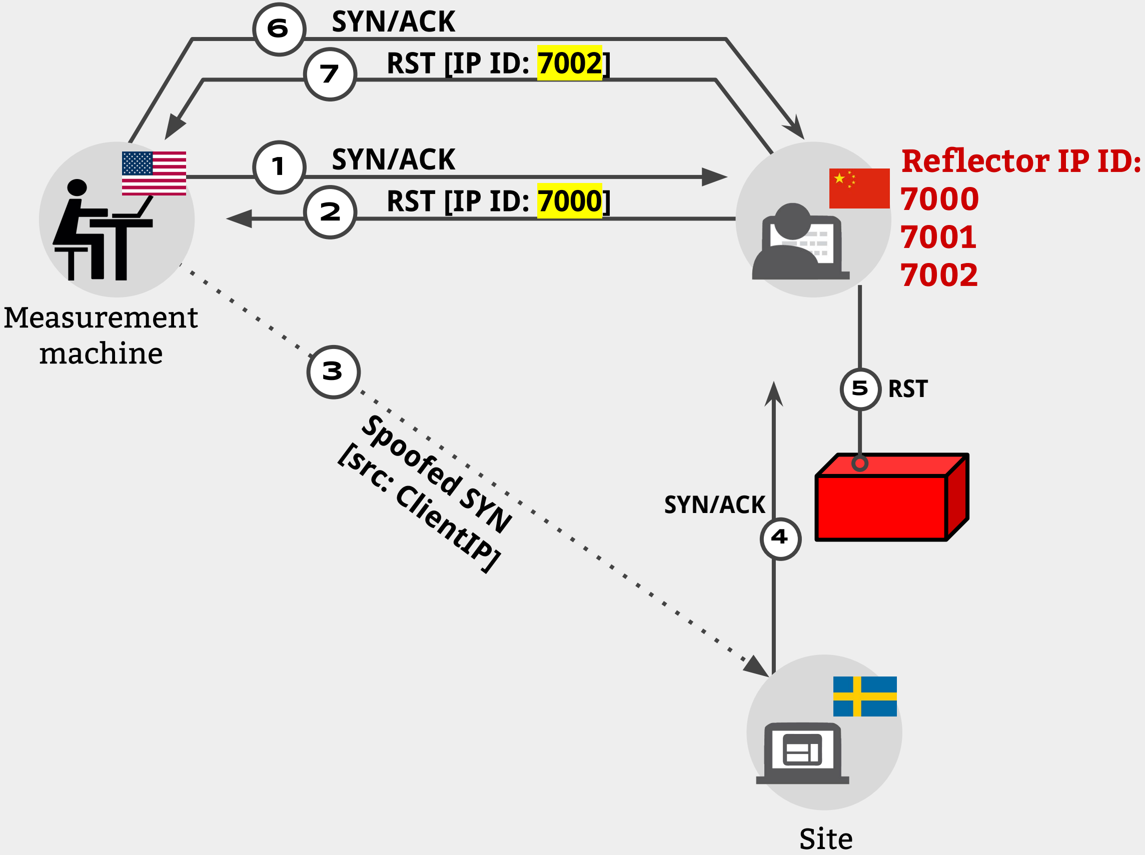
Spooky Scan

Site-to-Reflector
Blocked



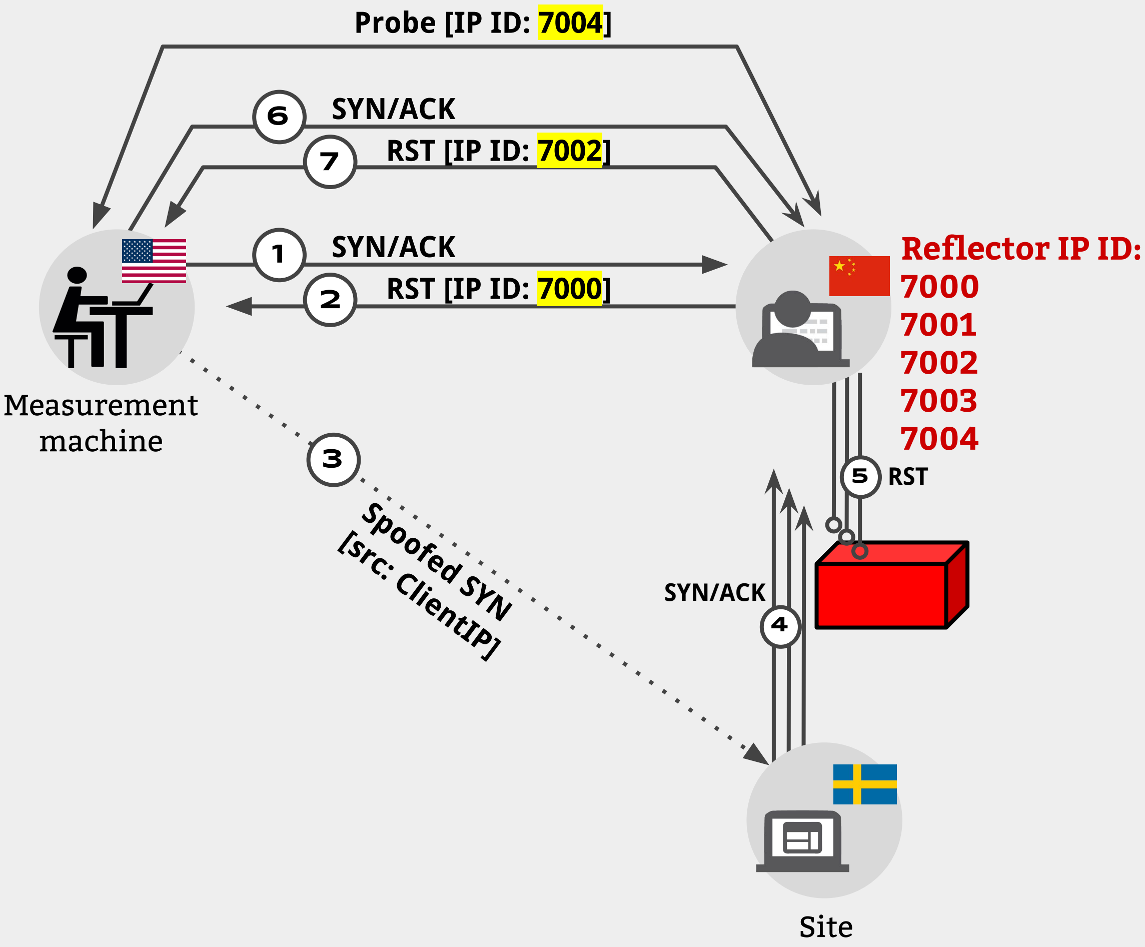
Spooky Scan

Reflector-to-Site Blocked



Spooky Scan

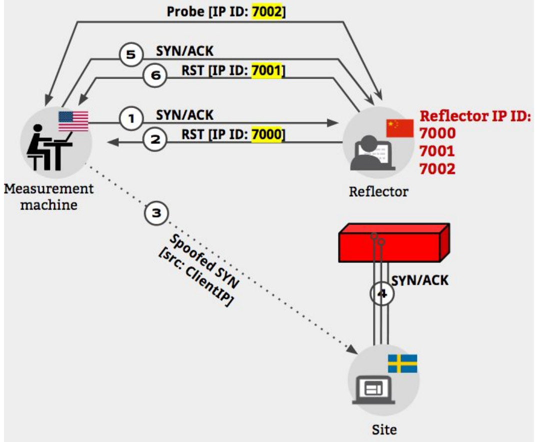
Reflector-to-Site Blocked



Spooky Scan

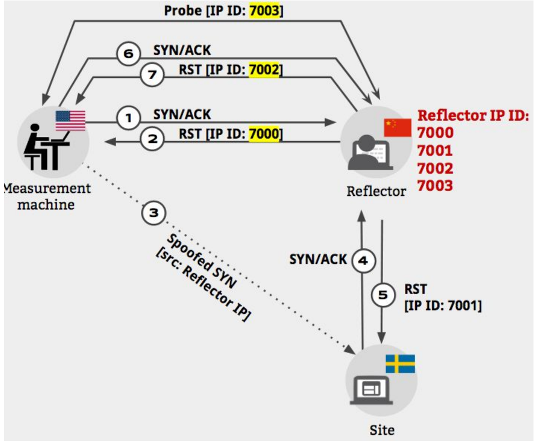
Site-to-Reflector Blocked

∆ IP ID1 = 1
∆ IP ID2 = 1



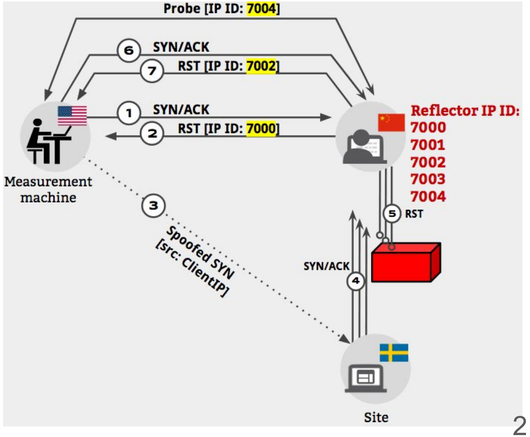
No Direction Blocked

∆ IP ID1 = 2
∆ IP ID2 = 1

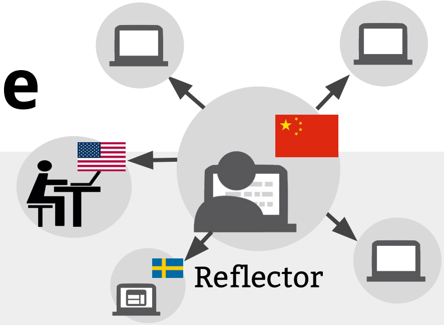


Reflector-to-Site Blocked

∆ IP ID1 = 2
∆ IP ID2 = 2



Coping with Reflector IP ID Noise

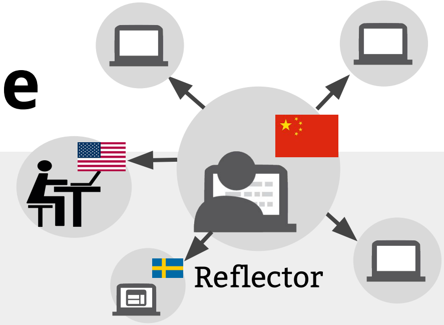


Amplifying the signal

Effect of sending *N* spoofed SYNs:

Site-to-Reflector Blocked	No Direction Blocked	Reflector-to-Site Blocked
Δ IP ID1 = (1 + noise) Δ IP ID2 = noise	Δ IP ID1 = (1 + N + noise) Δ IP ID2 = noise	Δ IP ID1 = (1 + N + noise) Δ IP ID2 = (1 + N + noise)

Coping with Reflector IP ID Noise



Amplifying the signal

Effect of sending *N* spoofed SYNs:

Site-to-Reflector Blocked	No Direction Blocked	Reflector-to-Site Blocked
Δ IP ID1 = (1 + noise) Δ IP ID2 = noise	Δ IP ID1 = (1 + N + noise) Δ IP ID2 = noise	Δ IP ID1 = (1 + N + noise) Δ IP ID2 = (1 + N + noise)

Repeating the experiment

To eliminate the effects of packet loss, sudden bursts of packets, ...

Augur for Continuous Scanning

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):

Run

- 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):

Run
- 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.

Augur: Sequential Hypothesis Testing

Defining a random variable:

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

Augur: Sequential Hypothesis Testing

Defining a random variable:

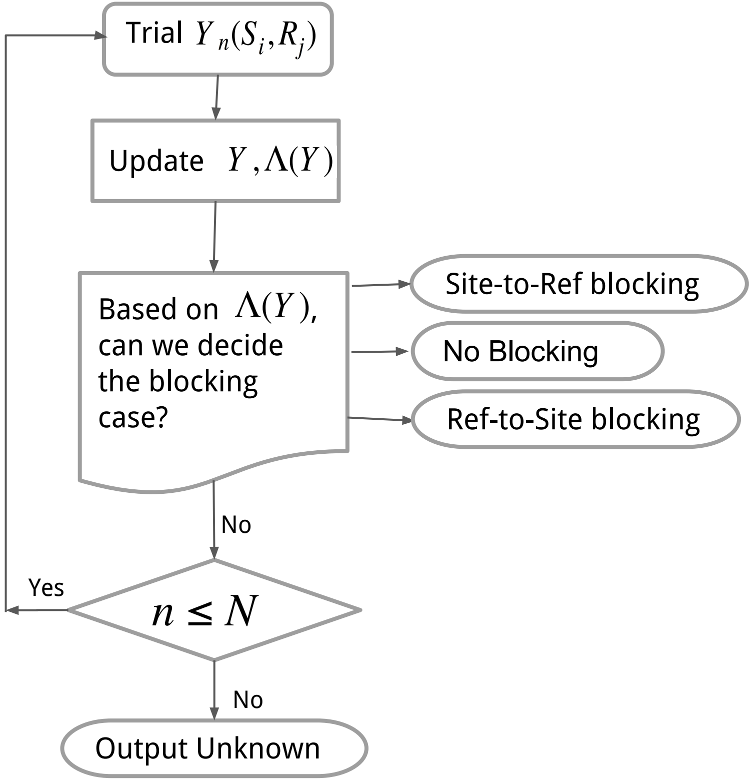
$$Y_n(S_i, R_j) = \begin{cases} 1 & \text{if no IPID acceleration occurs} \\ 0 & \text{if IPID acceleration occurs} \end{cases}$$

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

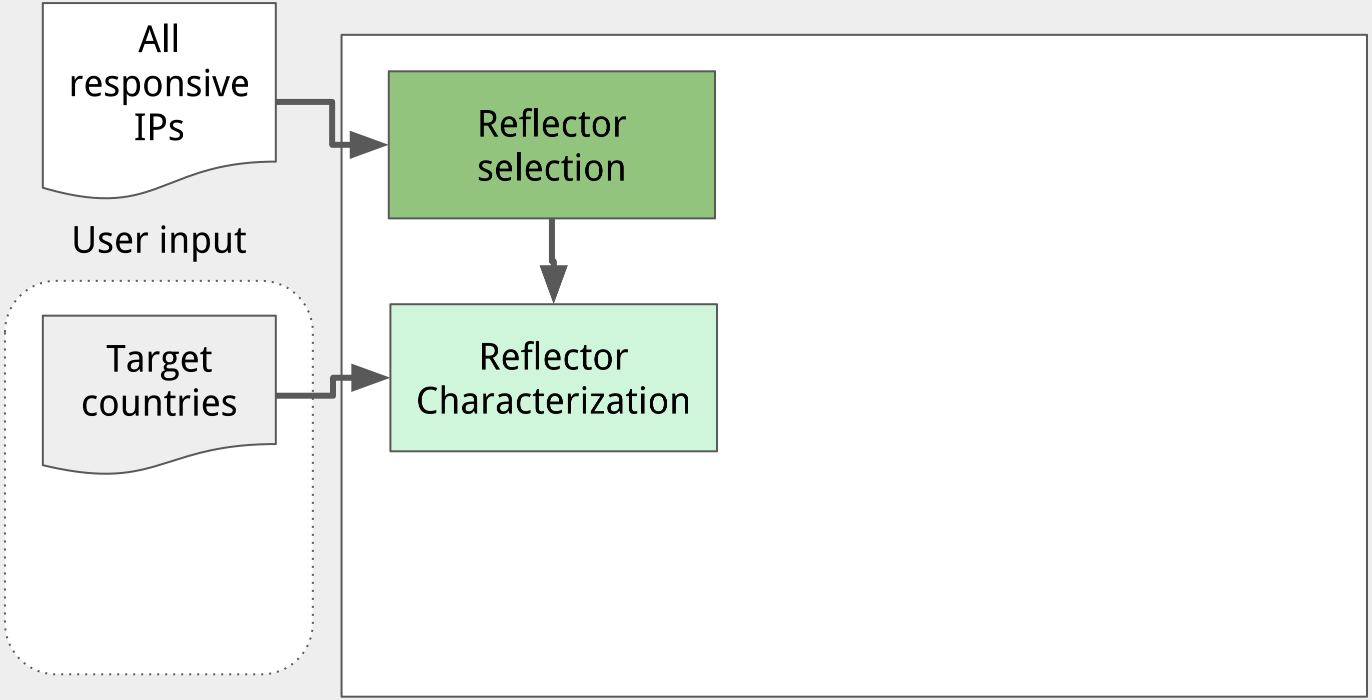


Maximum Likelihood Ratio

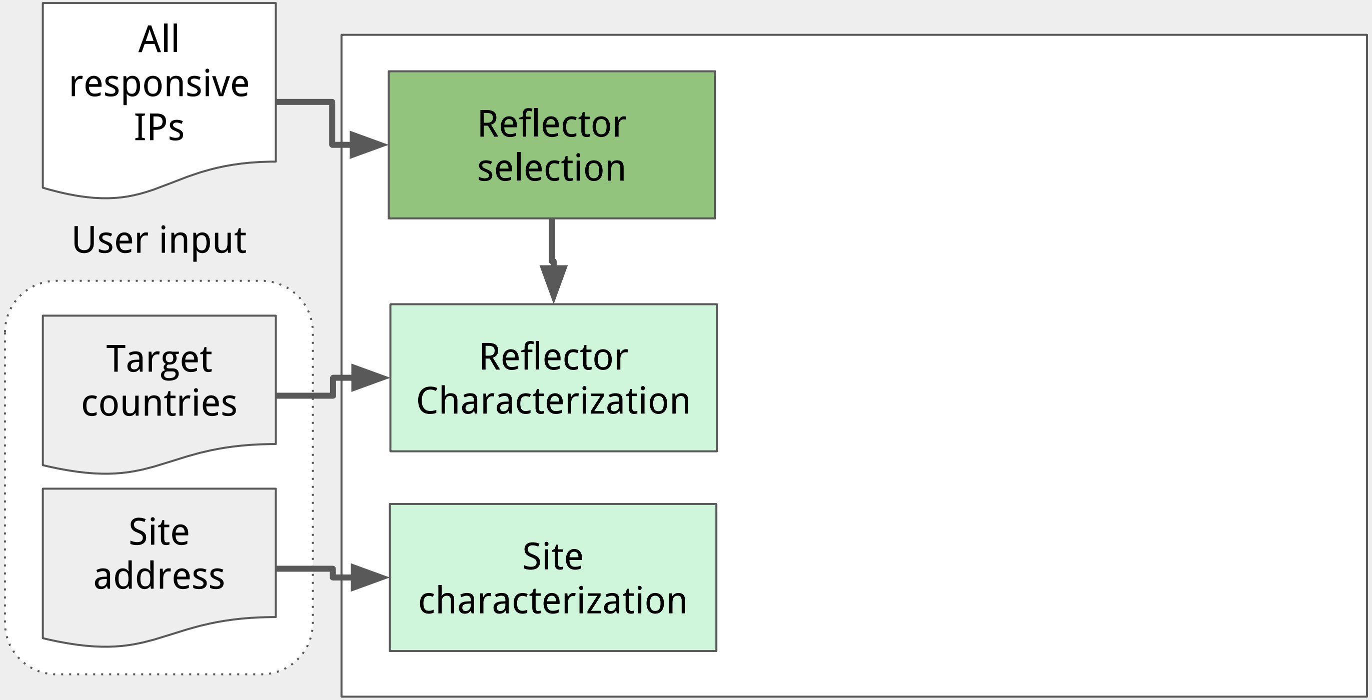
$$\Lambda(Y) \equiv \prod_{n=1}^{N} \frac{Pr[Y_n | Blocking]}{Pr[Y_n | No\ Blocking]}$$

Augur Framework

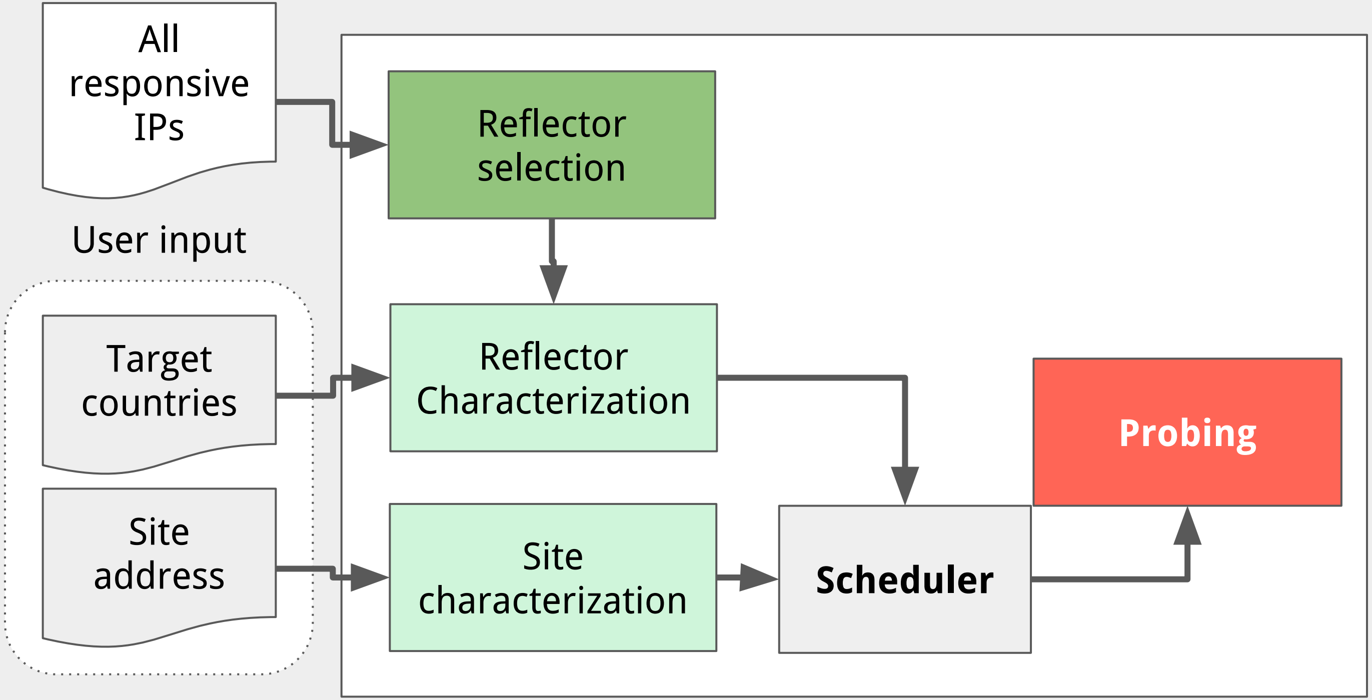
Augur Framework



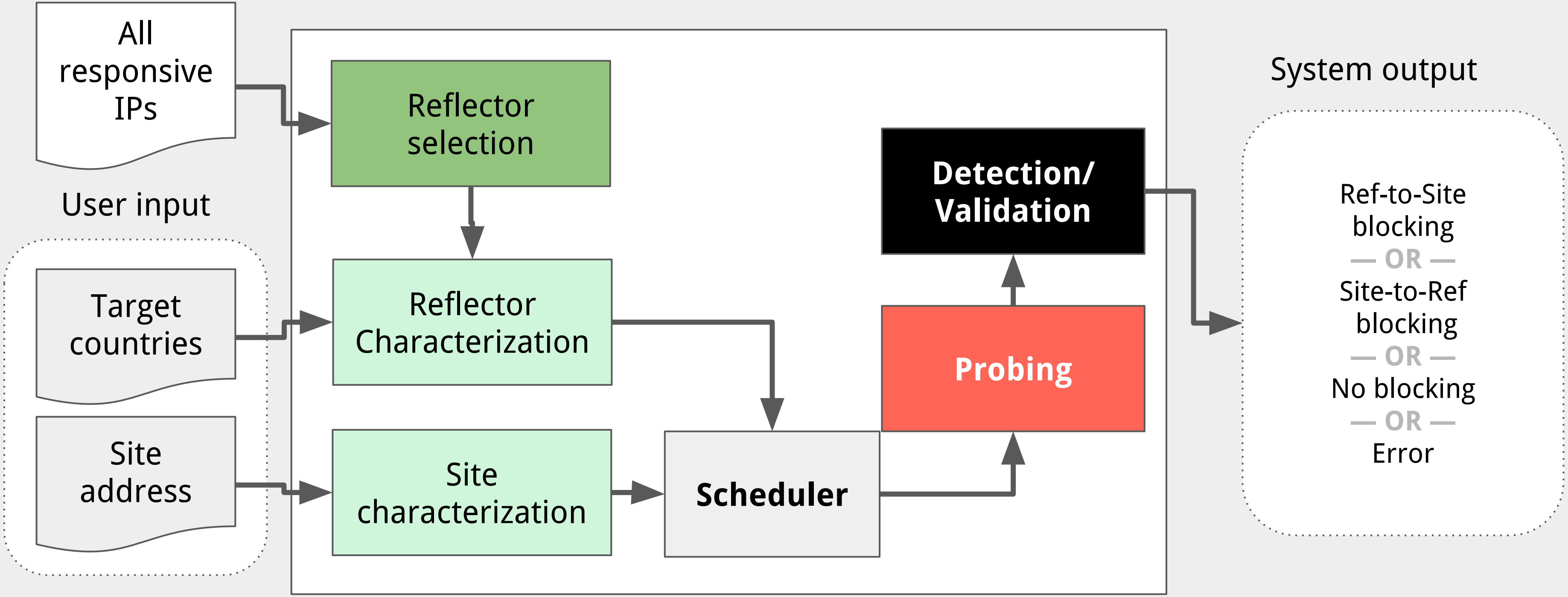
Augur Framework



Augur Framework



Augur Framework



Coverage

Challenge: Need global vantage points from which to measure

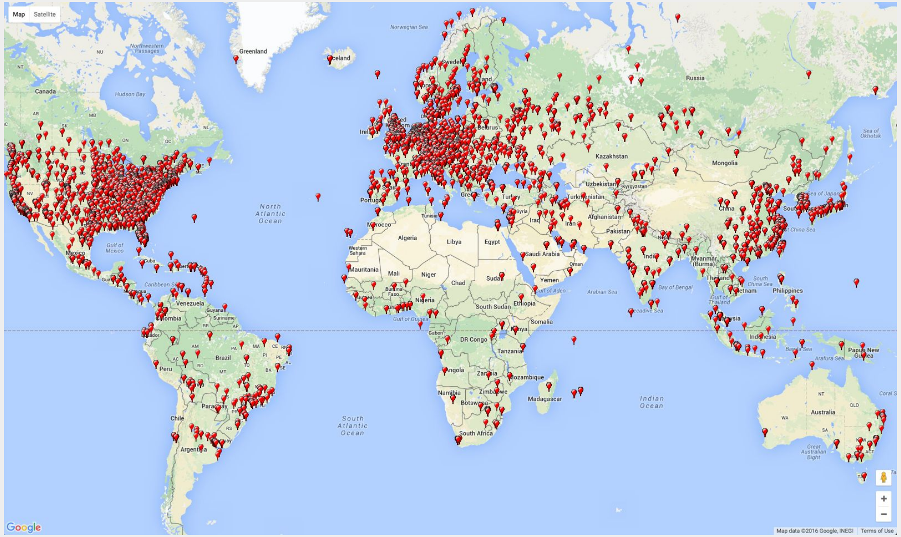
THREE KEY CHALLENGES:

✓ Coverage, ethics, and continuity

Scanning IPv4 on port 80:

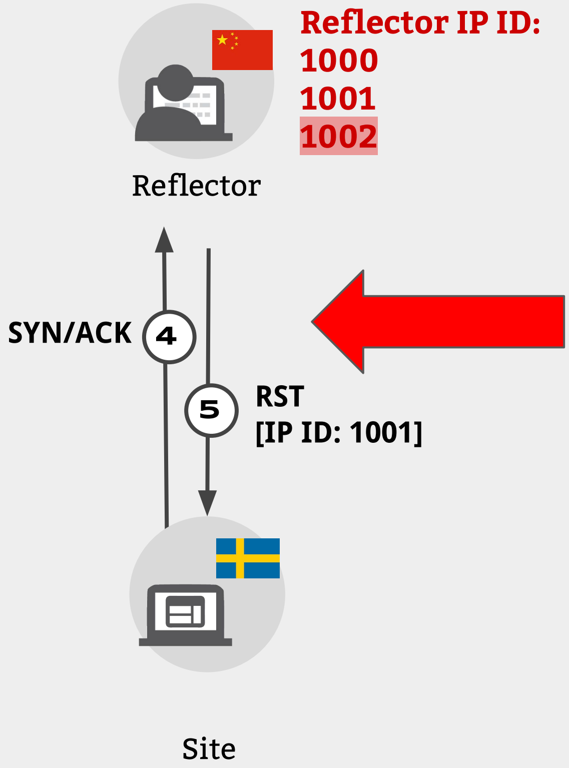
- 22.7 million potential reflectors!

Compare: 10,000 in prior work (RIPE Atlas)



Ethics

Challenge: Probing banned sites from users' machines creates risk

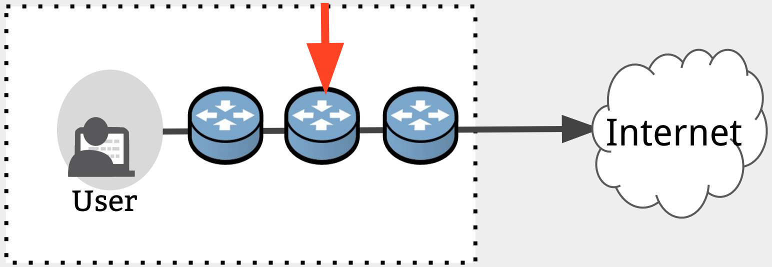


THREE KEY CHALLENGES:
Coverage ✓, ethics ✓, and continuity

Ethics

Challenge: Probing banned sites from users' machines creates risk

Use only **infrastructure devices** to source probes



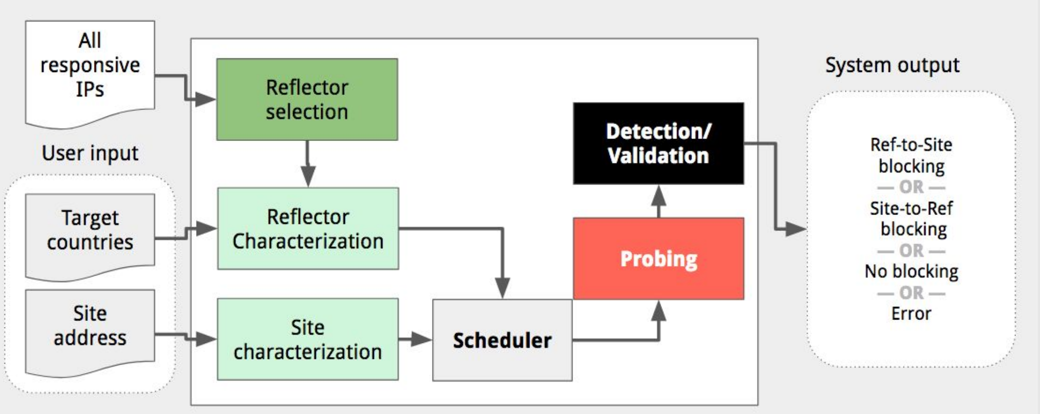
Global IP ID	22.7 million	236 countries (and dependent territories)
Two hops back from end user	**53,000**	**180 countries**

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

Continuity

Challenge: Need to repeat measurements over time

Augur doesn't depend on end users' availability, and routers have less downtime, allowing us to collect measurements continuously.



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

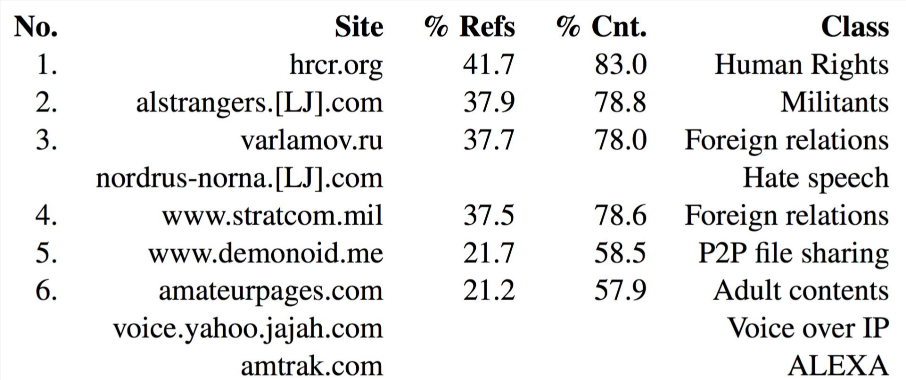
Measurements per reflector-site: 47

Overall # of measurements: 207.6 million

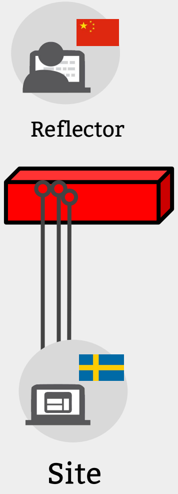
Top Blocked Sites

Site-to-Reflector Blocked

Site-to-Reflector blocking



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



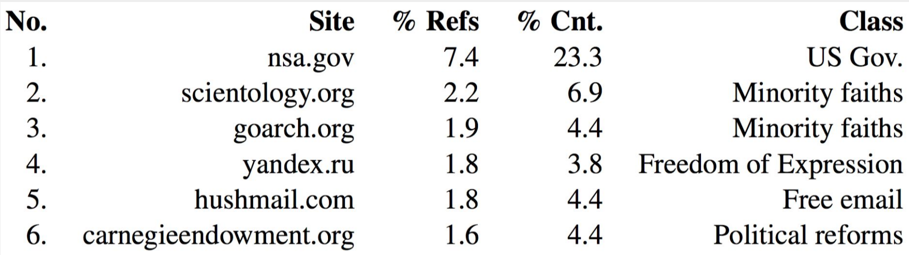
Interesting example:

- amtrak.com was blocked for 21% of reflectors, 57% of countries (ranked 6) → Collateral damage

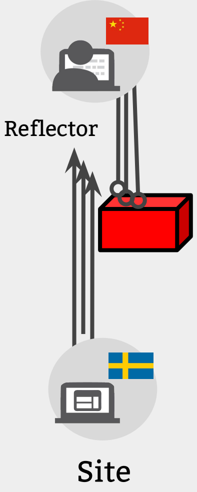
Top Blocked Sites

Reflector-to-site Blocked

Reflector-to-site blocking



No.	Site	% Refs	% Cnt.	Class
1.	nsa.gov	7.4	23.3	US Gov.
2.	scientology.org	2.2	6.9	Minority faiths
3.	goarch.org	1.9	4.4	Minority faiths
4.	yandex.ru	1.8	3.8	Freedom of Expression
5.	hushmail.com	1.8	4.4	Free email
6.	carnegieendowment.org	1.6	4.4	Political reforms



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

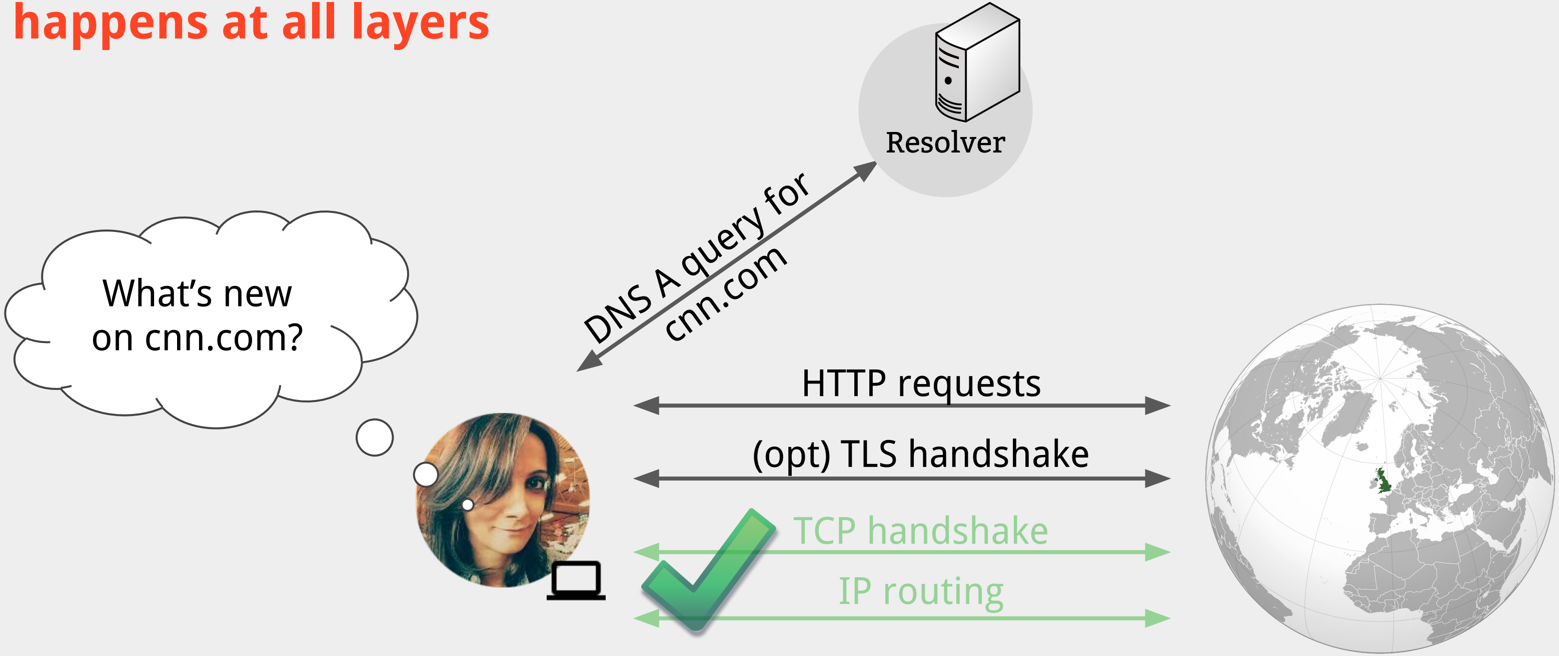
Augur

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

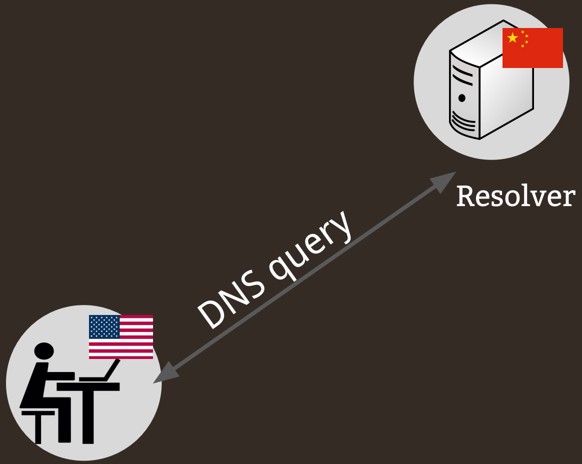
Network interference happens at all layers



Satellite (Iris)

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**



* **Satellite: Joint Analysis of CDNs and Network-Level Interference,Satelite,** Scott, Anderson, Kohno, and Krishnamurthy. In USENIX ATC, 2016.
* **Global Measurement of DNS Manipulation,** Pearce, Jones, Li, Ensafi , Feamster, Paxson, USENIX Security, August 2017

Deploying Satellite

Challenge:
Identify “wrong”
DNS responses

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

Coverage:

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

Ethical:

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

Continuity:

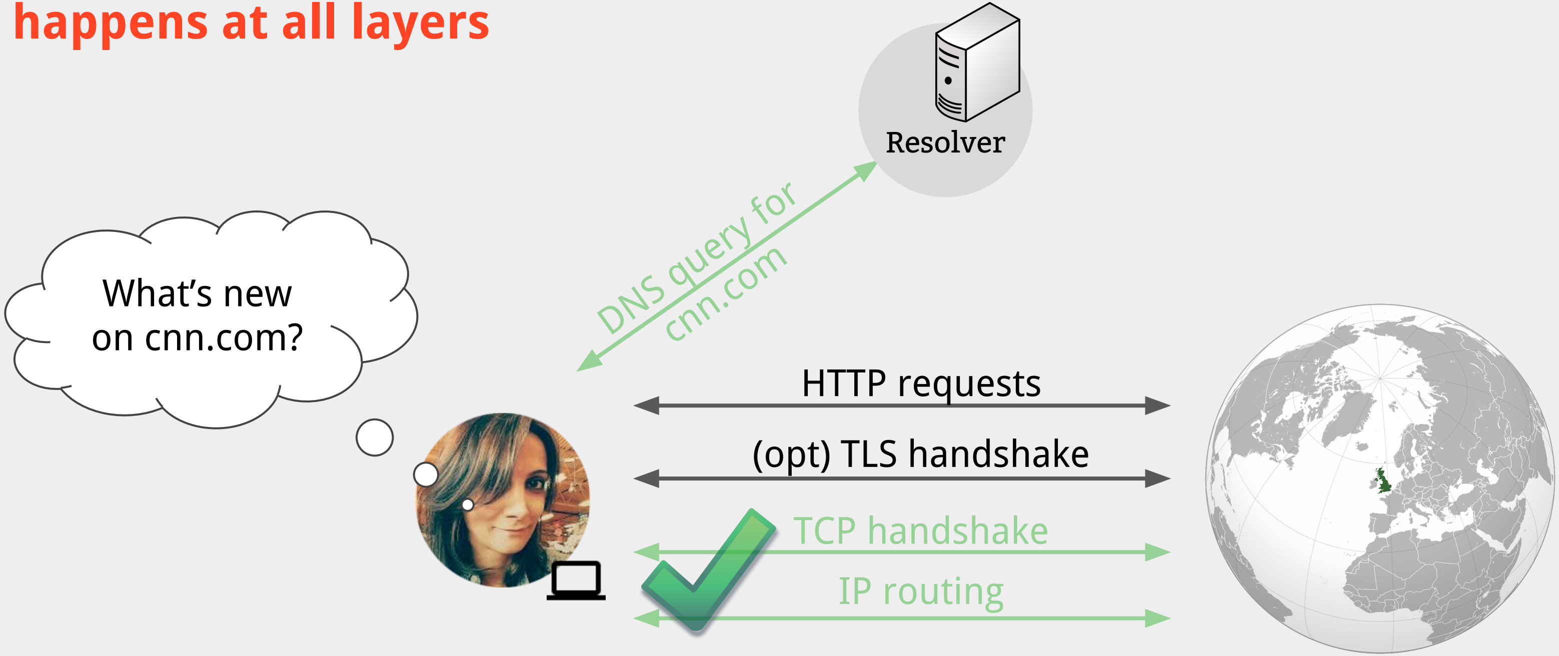
- Satellite doesn’t depend on end users’ availability, and resolvers have less downtime

Detecting DNS manipulation:

- Using consistency and independent verifiability heuristics.

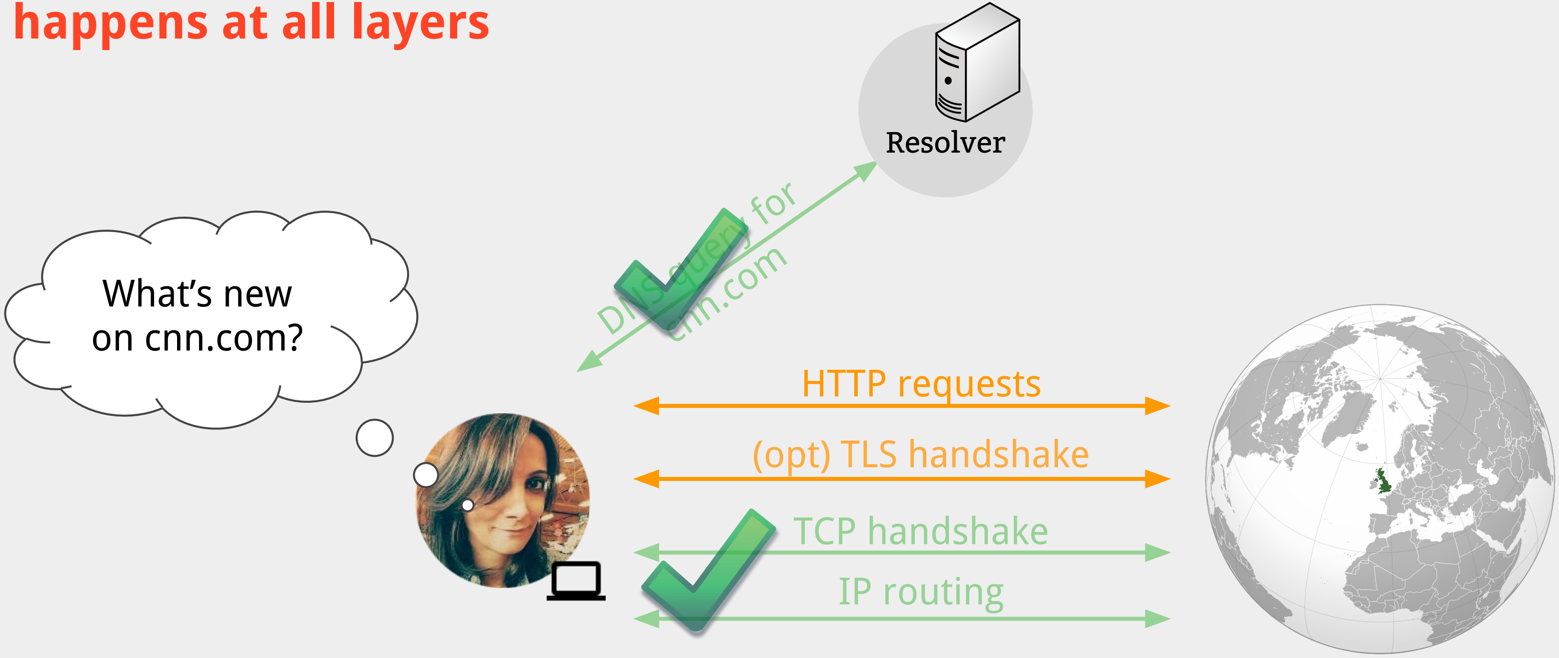
Side Channels at Other Network Layers

Network interference happens at all layers



Side Channels at Other Network Layers

Network interference happens at all layers



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