Spoki: Unveiling a New Wave of Scanners through a Reactive Network Telescope

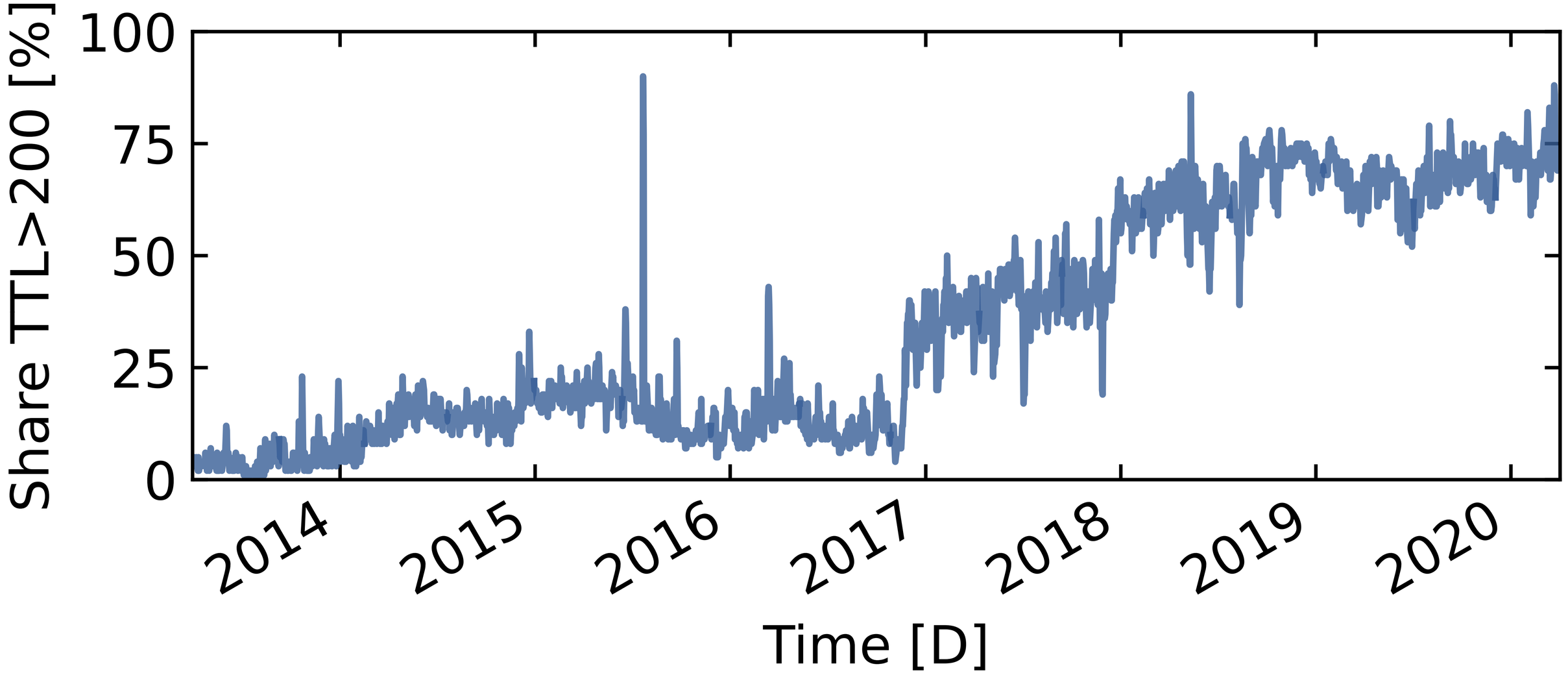
Raphael Hiesgen, Marcin Nawrocki, Alistair King,
Alberto Dainotti, Thomas C. Schmidt, Matthias Wählisch

USENIX Security, August 12, 2022

raphael.hiesgen@haw-hamburg.de

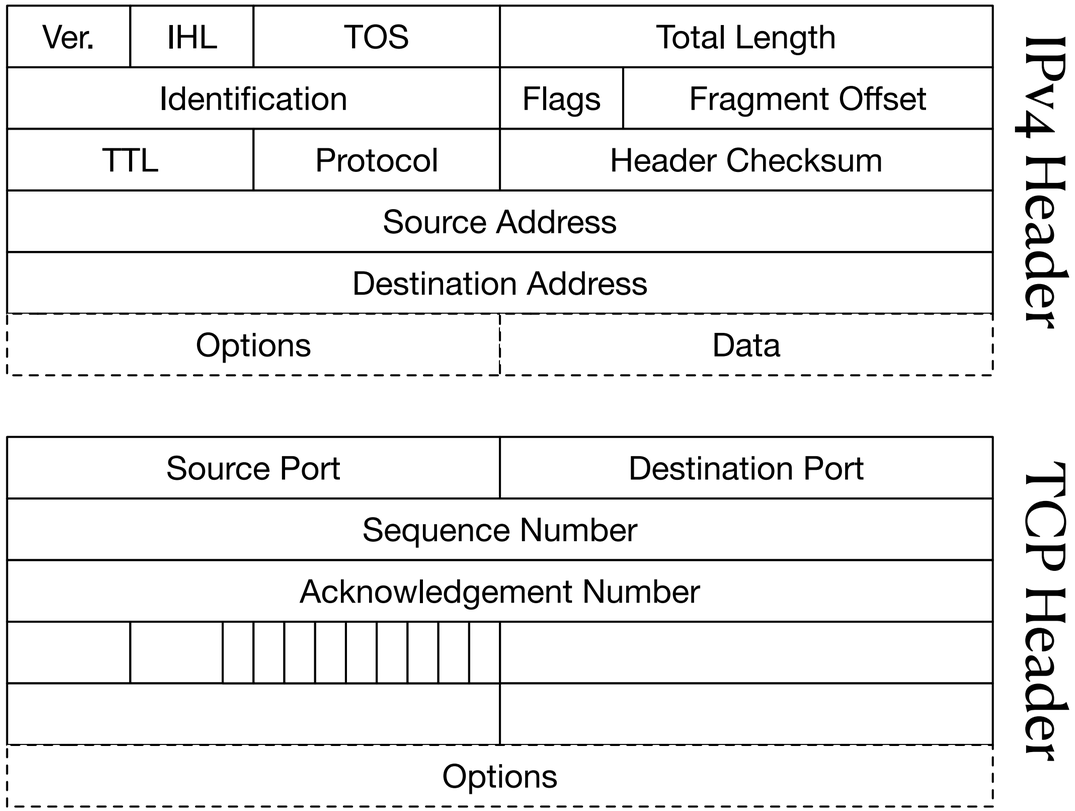
The Share of Irregular Packets is Increasing

UCSD Network Telescope: a /9 IPv4 prefix



What is a SYN Irregularity?

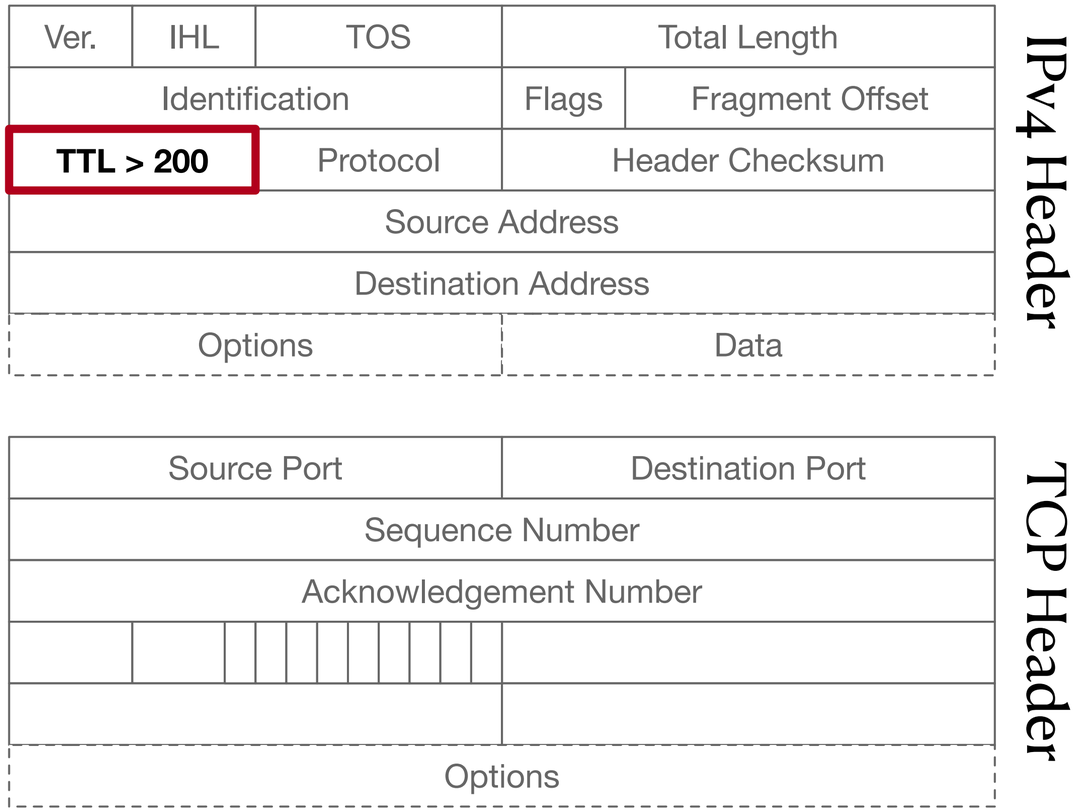
- Irregular packets show one or more of:
  - High TTL (≥200)
  - No TCP options
  - Fixed IP ID (54321)



- The telescope now observes a share of roughly 75% irregular SYNs

What is a SYN Irregularity?

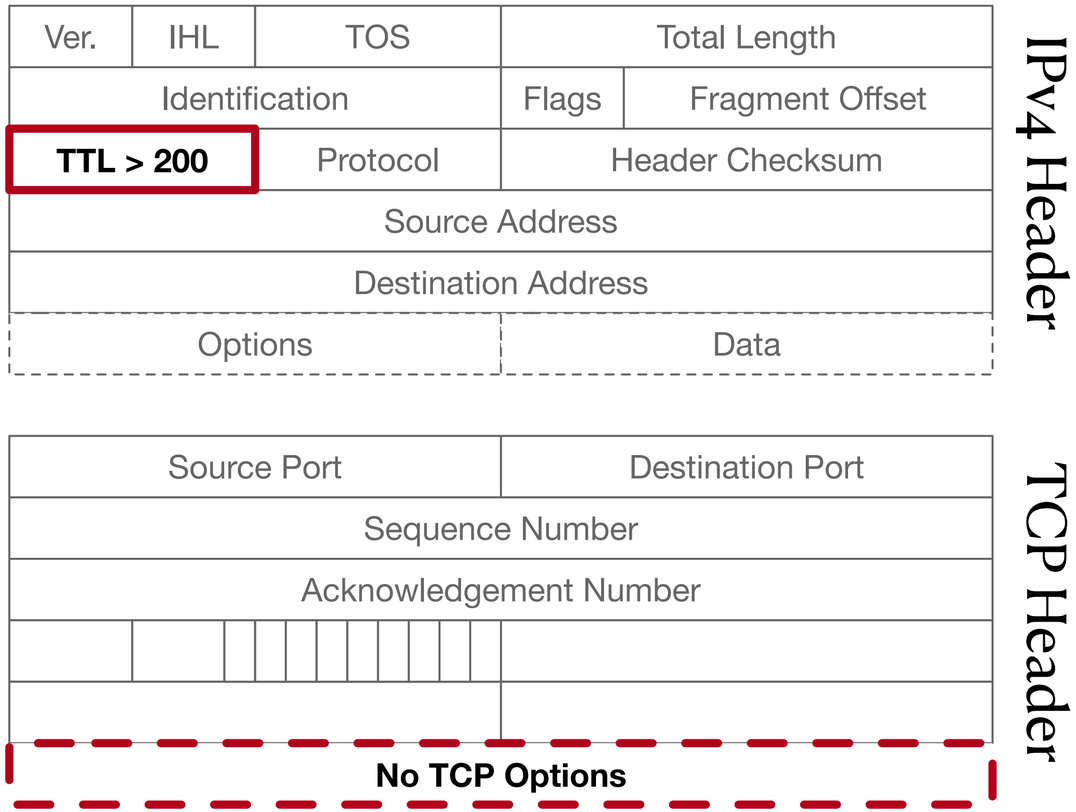
- Irregular packets show one or more of:
  - High TTL (≥200)
  - No TCP options
  - Fixed IP ID (54321)



- The telescope now observes a share of roughly 75% irregular SYNs

What is a SYN Irregularity?

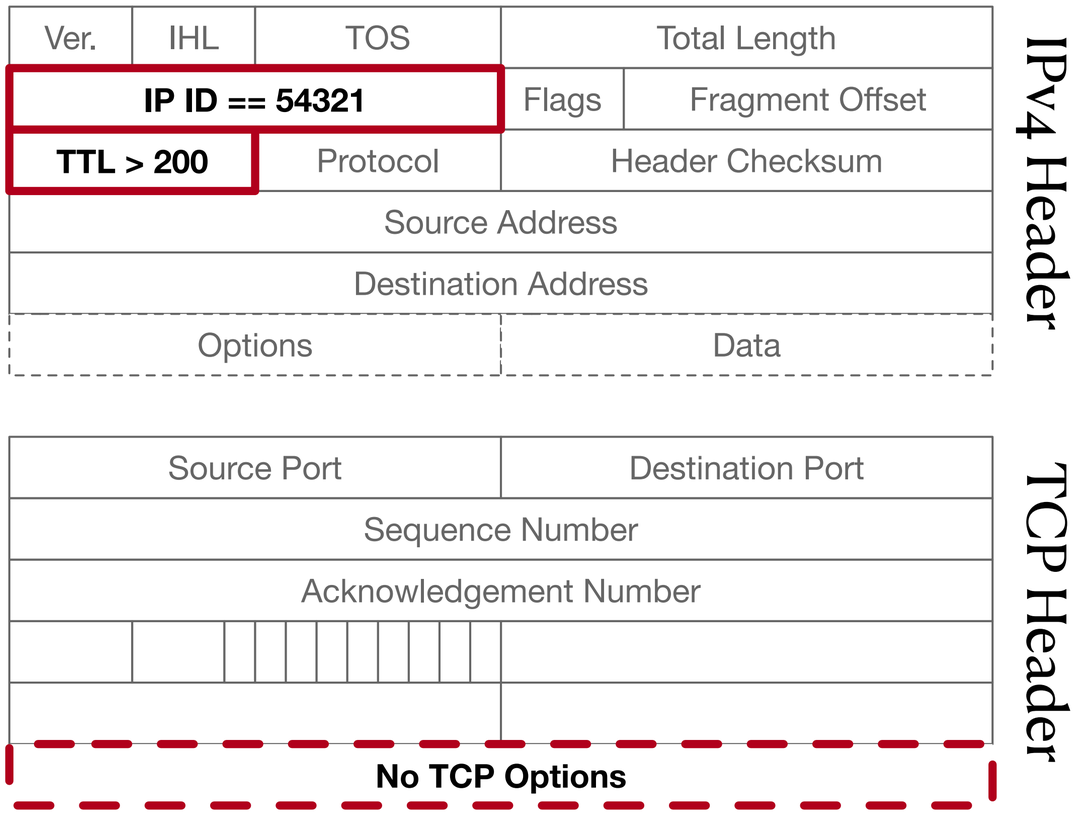
- Irregular packets show one or more of:
  - High TTL (≥200)
  - No TCP options
  - Fixed IP ID (54321)



- The telescope now observes a share of roughly 75% irregular SYNs

What is a SYN Irregularity?

- Irregular packets show one or more of:
  - High TTL (≥200)
  - No TCP options
  - Fixed IP ID (54321)



- The telescope now observes a share of roughly 75% irregular SYNs

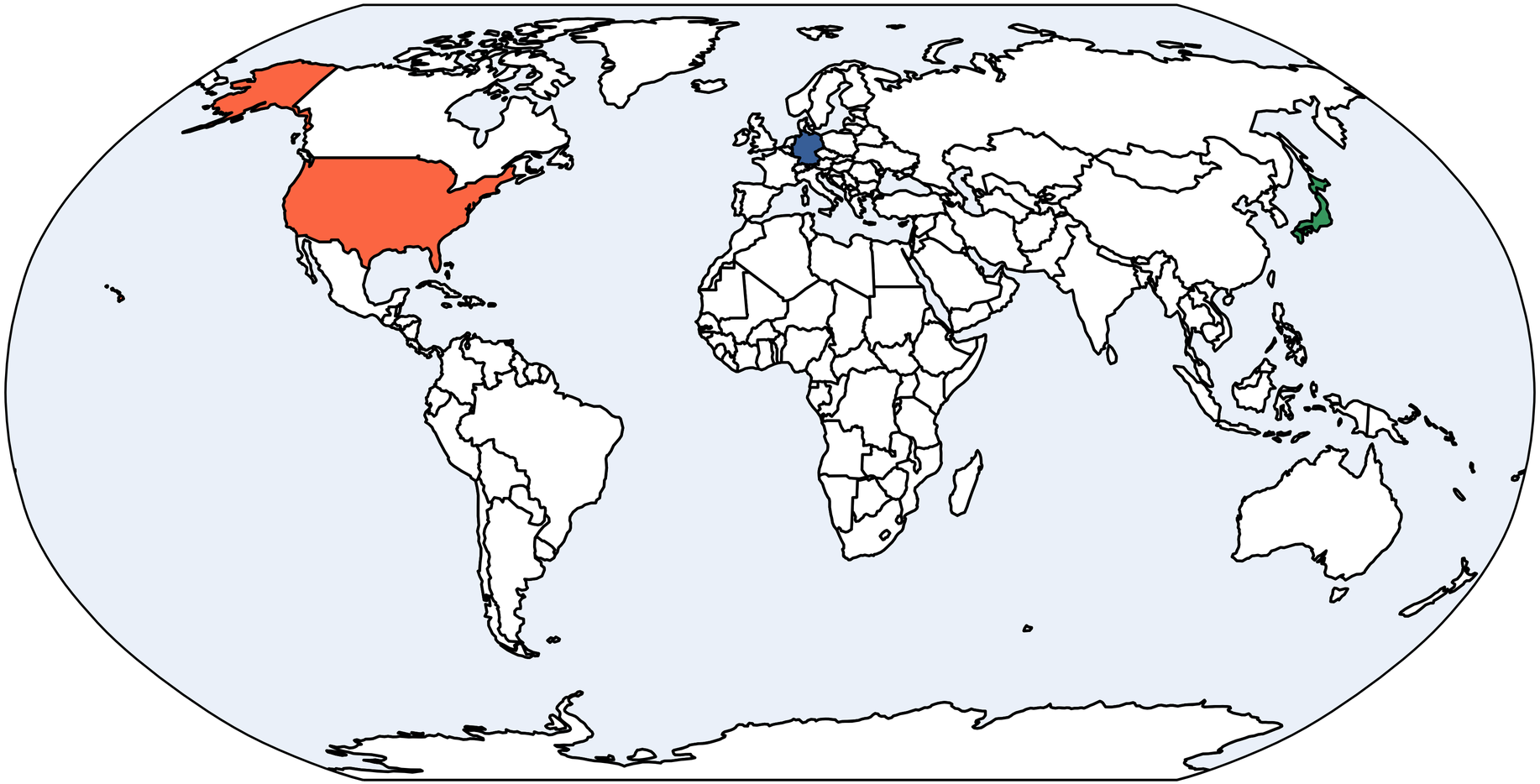
What is a SYN Irregularity?

- I
  -
  -
  -

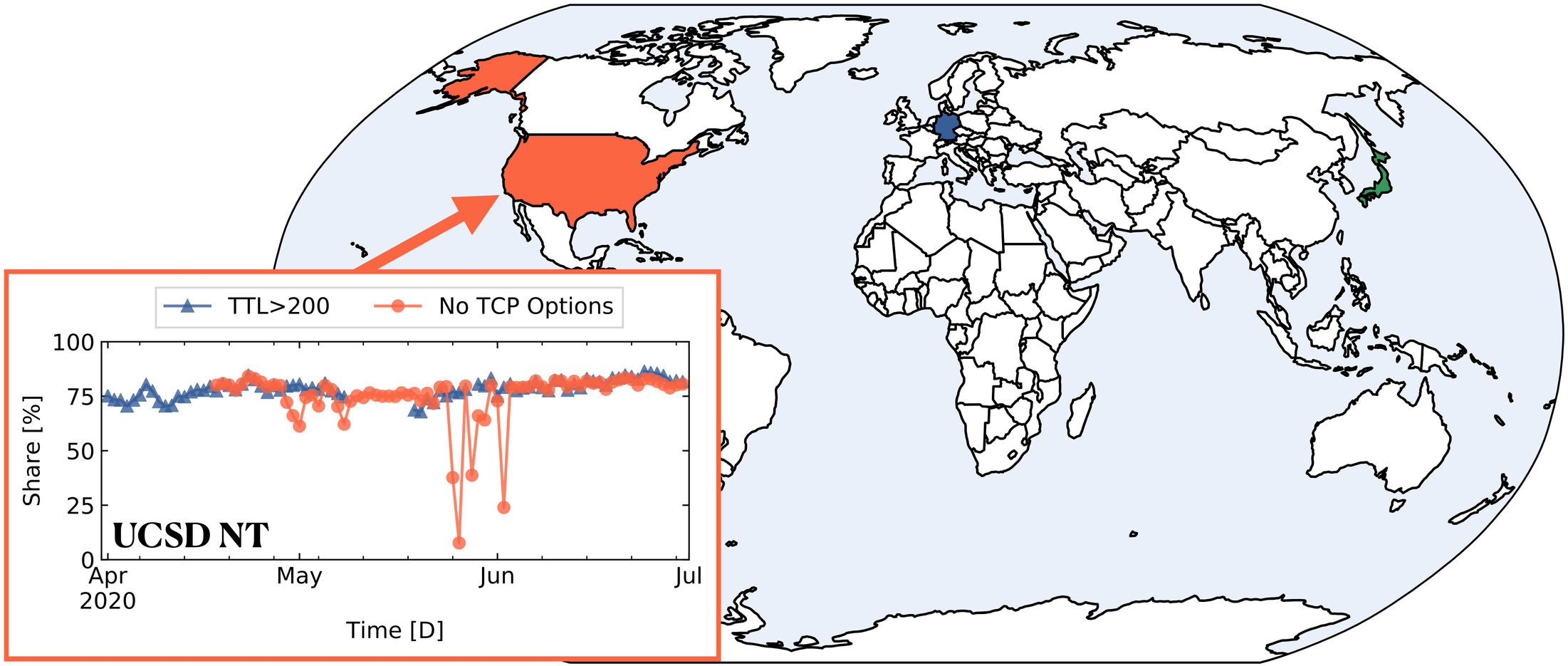
Is this observation specific to the UCSD network telescope?

- The telescope now observes a share of roughly 75% irregular SYNs

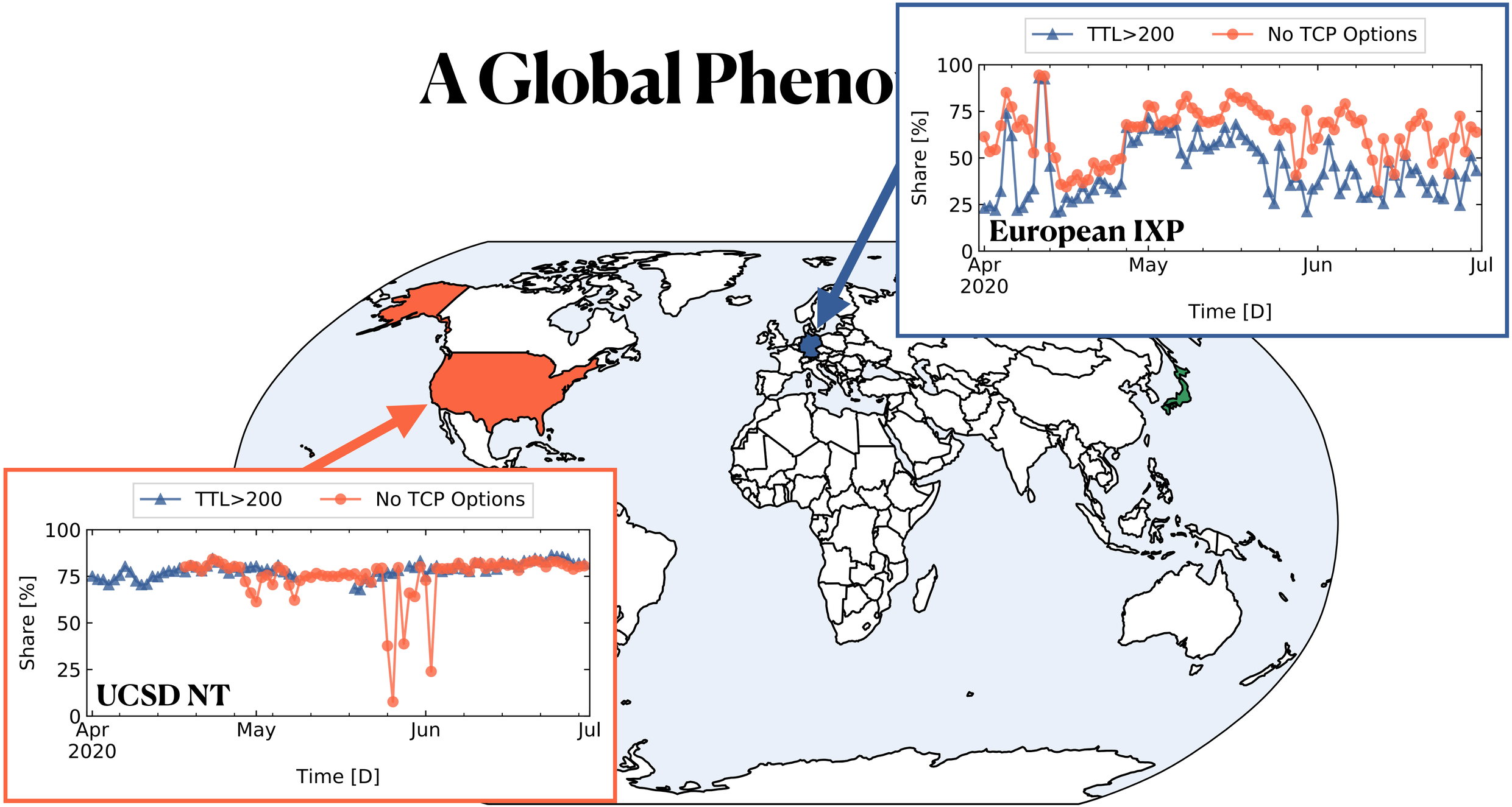
A Global Phenomenon



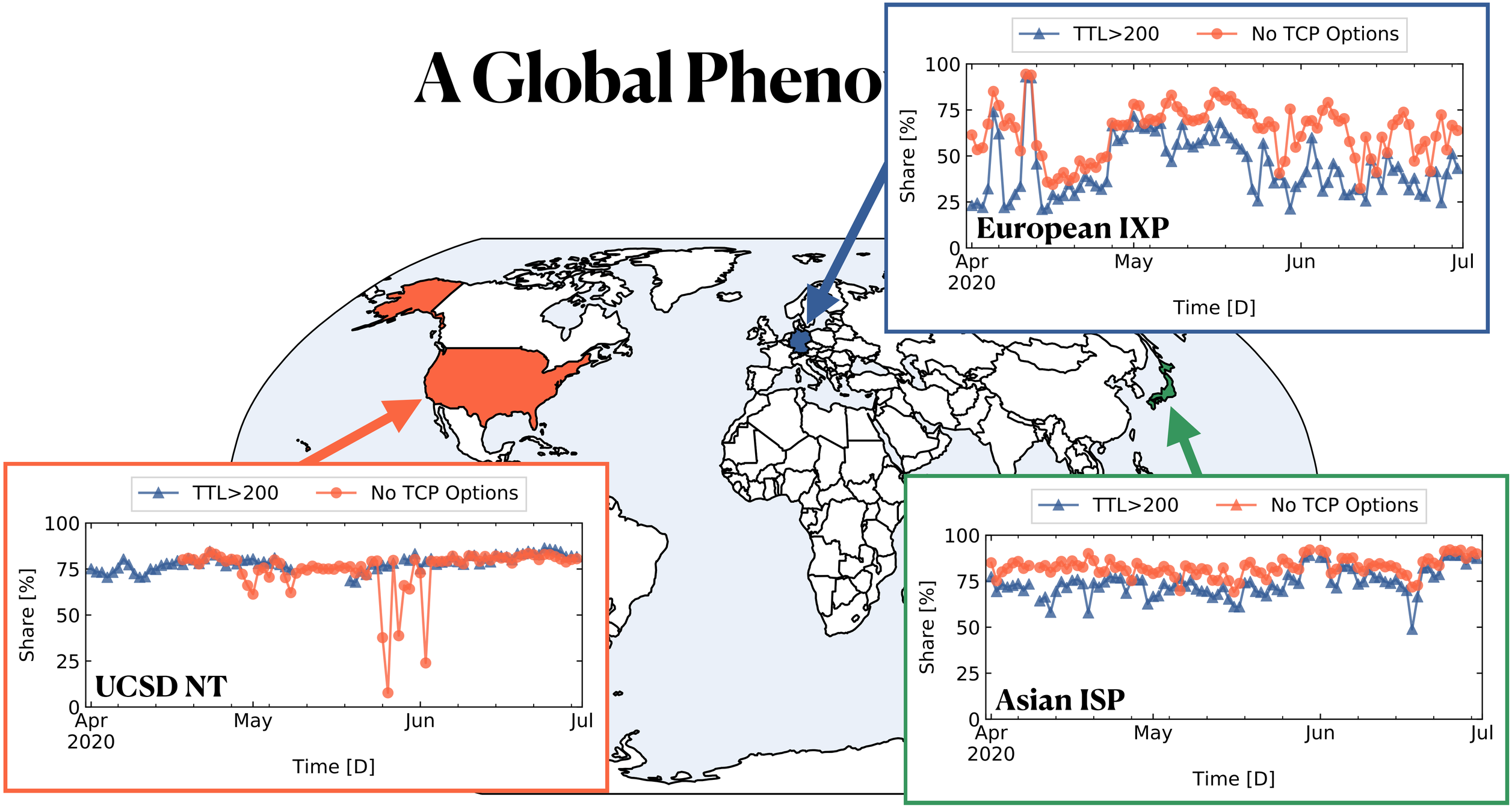
A Global Phenomenon



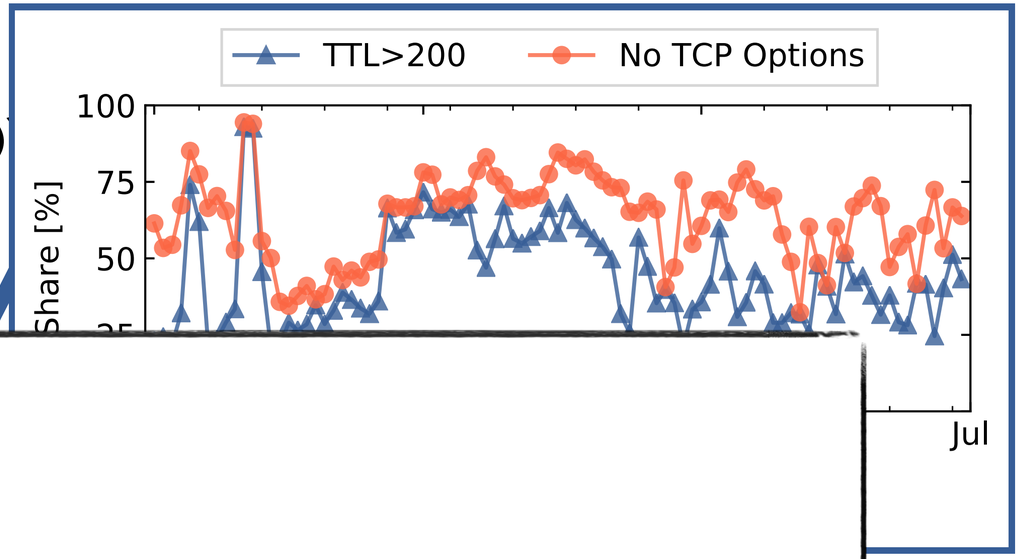
A Global Pheno



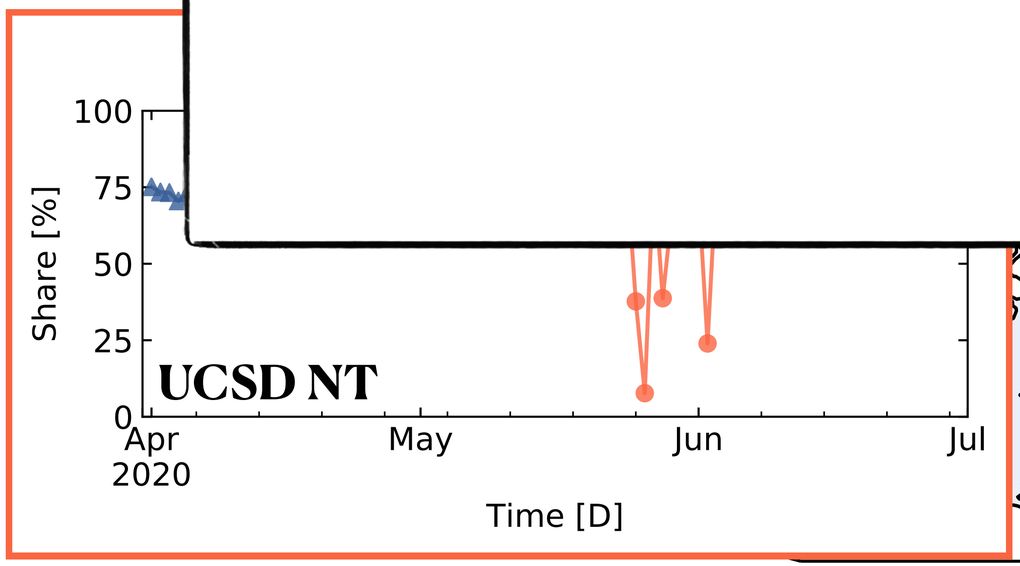
A Global Pheno

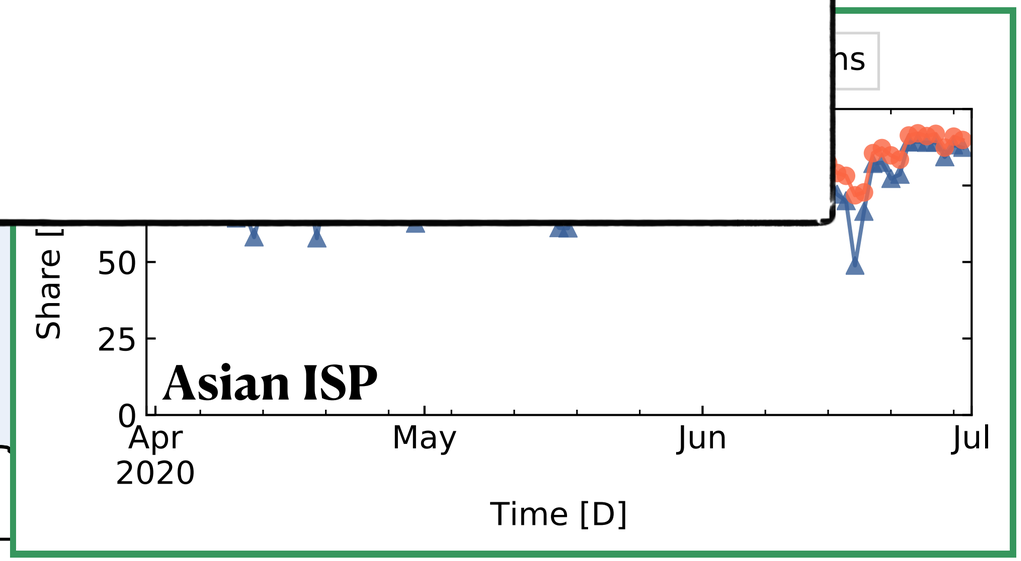


A Global Pheno



Do these packets pose a threat?





Background: Stateless Scanning

"Scan the Internet in less than 1 hour on commodity hardware!"

- Increases scan speeds by avoiding local state
  - Hand-crafted probes sent via raw sockets
  - Recognize replies via SYN cookies

- Popularized by **ZMap** around 2013
- Abused by **Mirai** in 2016

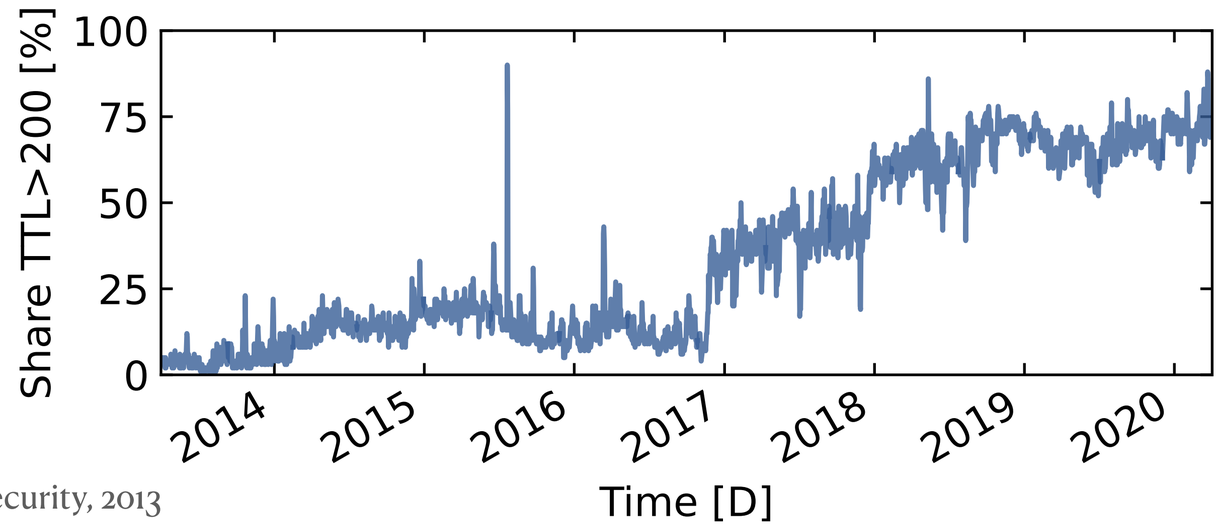
Durumeric et al., *ZMap: Fast Internet-Wide Scanning and its Security Applications*, USENIX Security, 2013
Antonakakis et al., *Understanding the Mirai Botnet*, USENIX Security, 2017

Background: Stateless Scanning

"Scan the Internet in less than 1 hour on commodity hardware!"

- Increases scan speeds by avoiding local state
  - Hand-crafted probes sent via raw sockets
  - Recognize replies via SYN cookies

- Popularized by **ZMap** around 2013
- Abused by **Mirai** in 2016



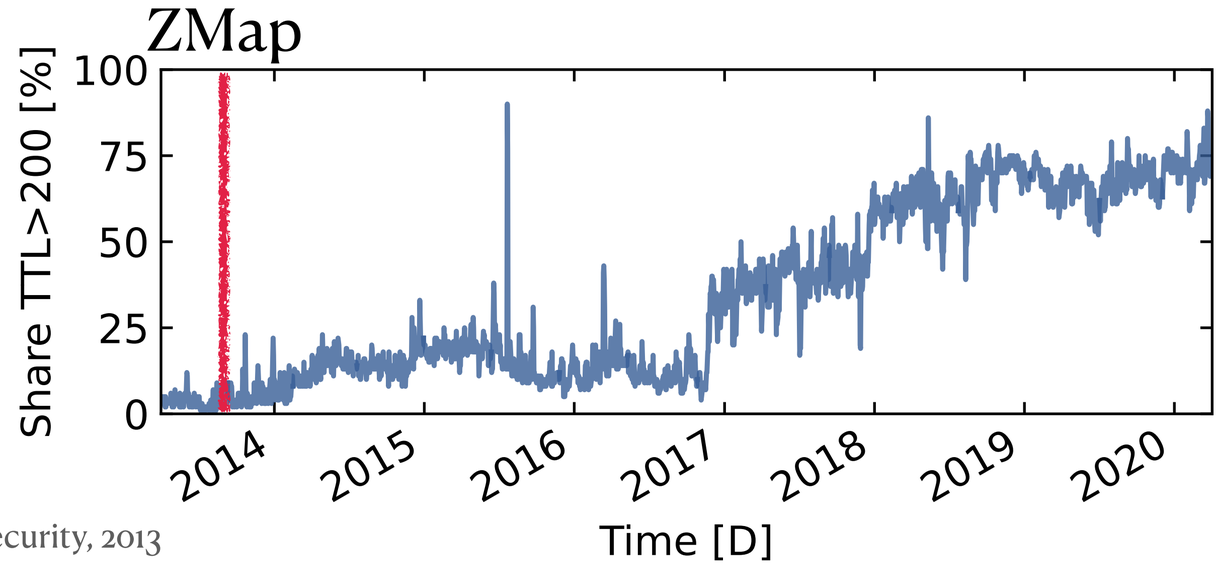
Durumeric et al., *ZMap: Fast Internet-Wide Scanning and its Security Applications*, USENIX Security, 2013
Antonakakis et al., *Understanding the Mirai Botnet*, USENIX Security, 2017

Background: Stateless Scanning

"Scan the Internet in less than 1 hour on commodity hardware!"

- Increases scan speeds by avoiding local state
  - Hand-crafted probes sent via raw sockets
  - Recognize replies via SYN cookies

- Popularized by **ZMap** around 2013
- Abused by **Mirai** in 2016



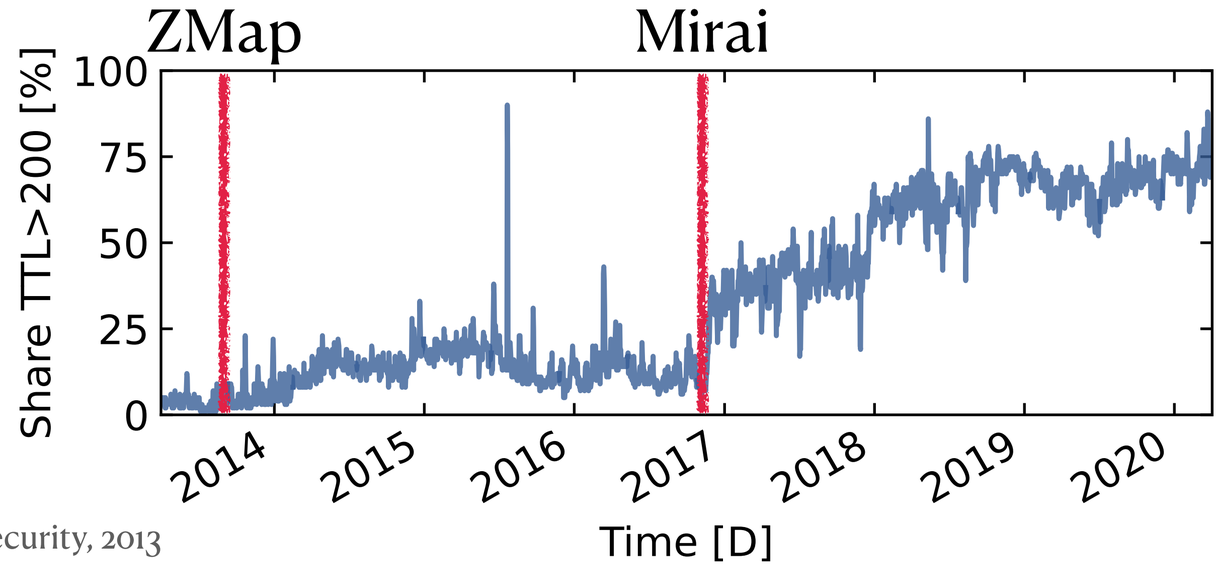
Durumeric et al., *ZMap: Fast Internet-Wide Scanning and its Security Applications*, USENIX Security, 2013
Antonakakis et al., *Understanding the Mirai Botnet*, USENIX Security, 2017

Background: Stateless Scanning

"Scan the Internet in less than 1 hour on commodity hardware!"

- Increases scan speeds by avoiding local state
  - Hand-crafted probes sent via raw sockets
  - Recognize replies via SYN cookies

- Popularized by **ZMap** around 2013
- Abused by **Mirai** in 2016



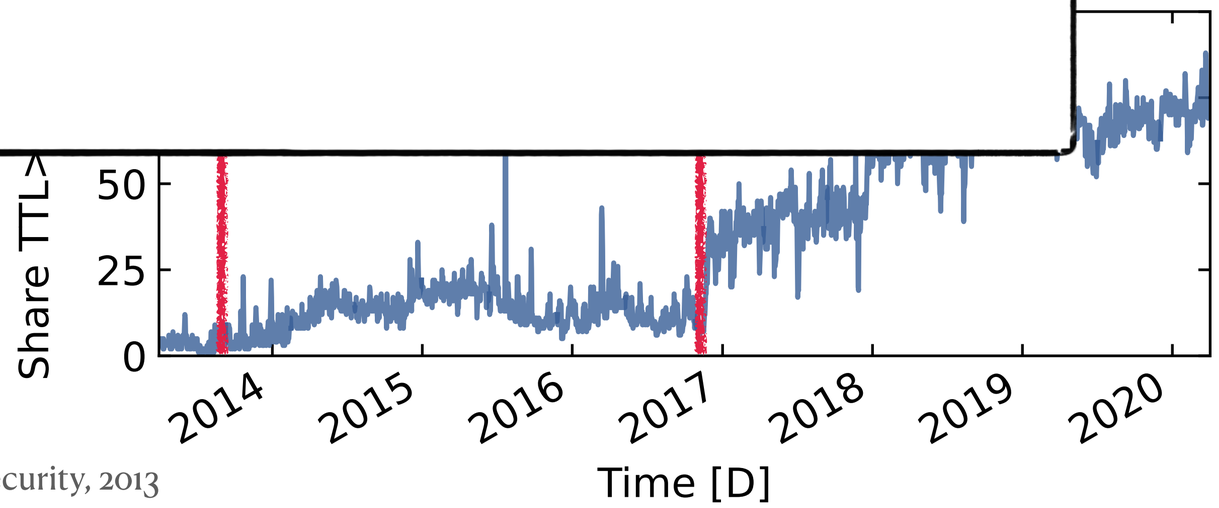
Durumeric et al., *ZMap: Fast Internet-Wide Scanning and its Security Applications*, USENIX Security, 2013
Antonakakis et al., *Understanding the Mirai Botnet*, USENIX Security, 2017

Background: Stateless Scanning

"Scan the Internet in less than 1 hour on commodity hardware!"

- I
  -
  -
- P
- A

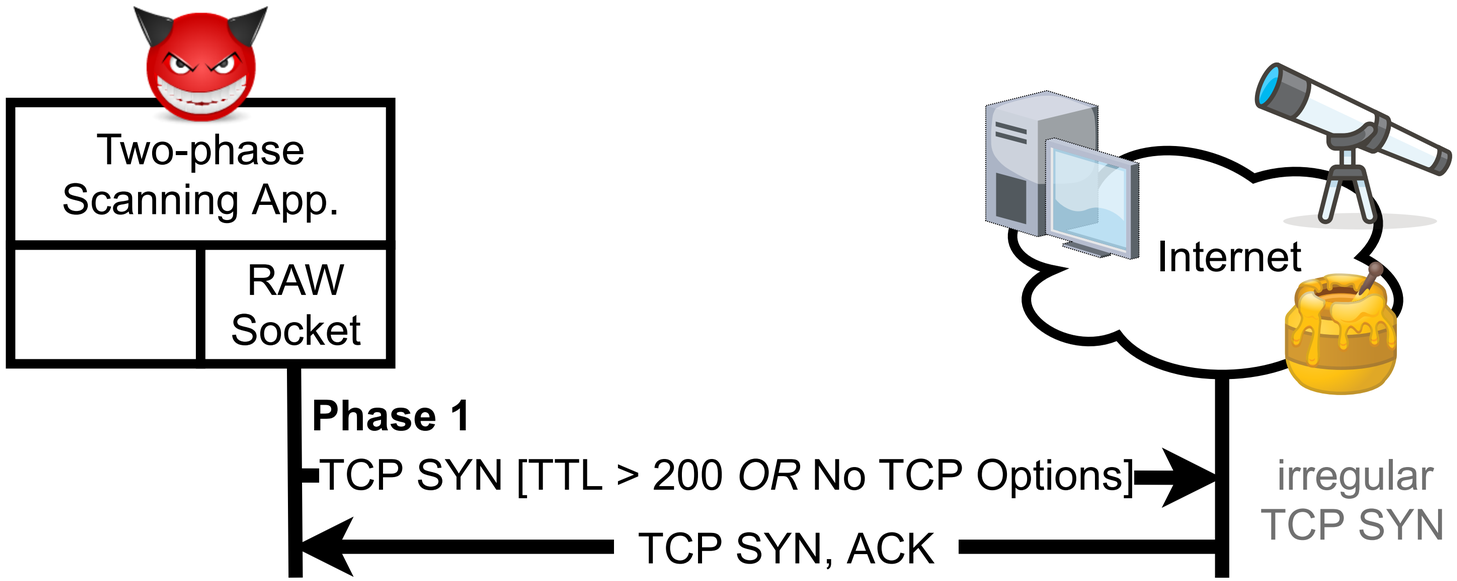
How can stateless scanning be abused?



Durumeric et al., *ZMap: Fast Internet-Wide Scanning and its Security Applications*, USENIX Security, 2013
Antonakakis et al., *Understanding the Mirai Botnet*, USENIX Security, 2017

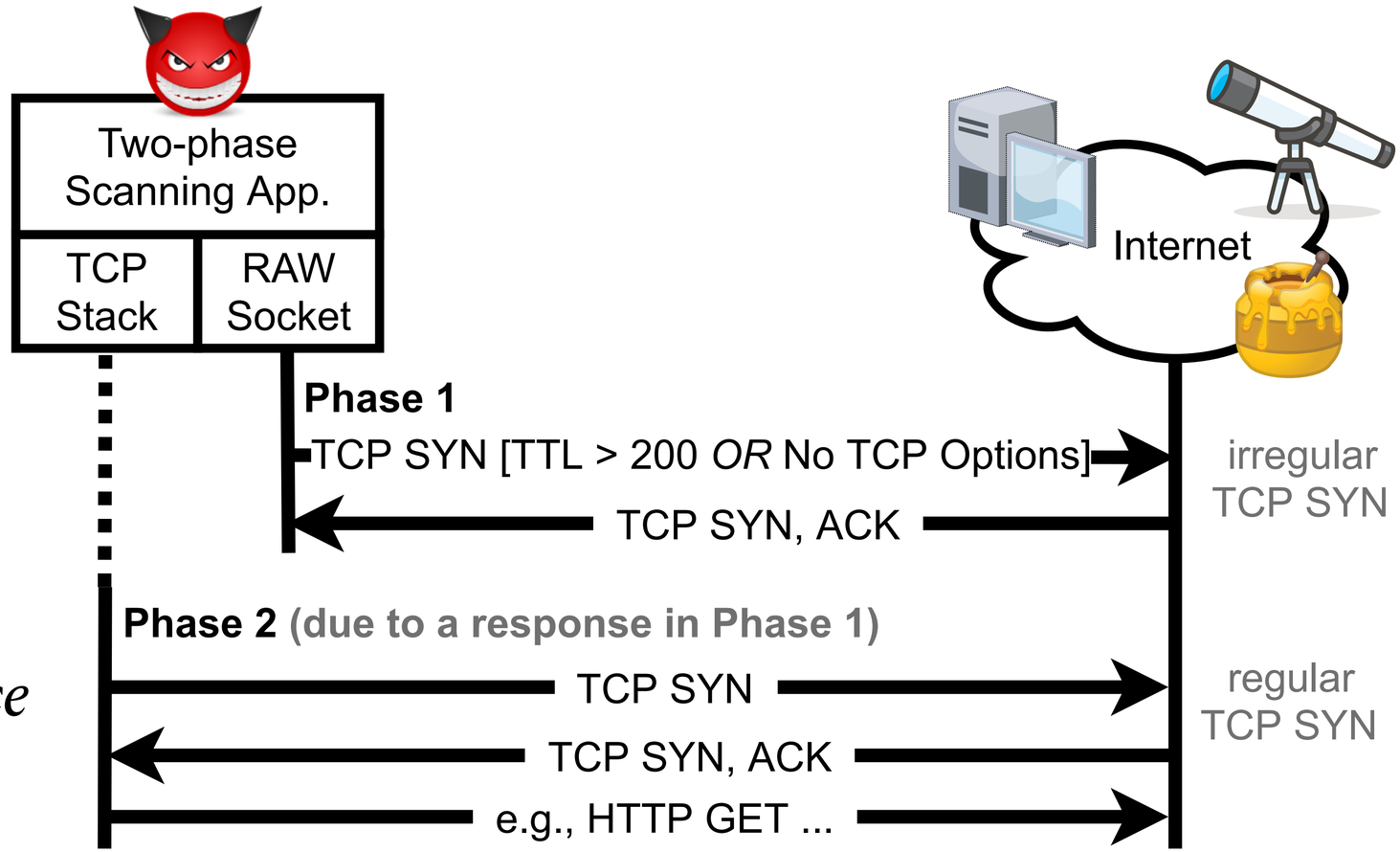
Two-phase Scanning

- First phase: Transport layer
  - Hand-crafted, stateless SYNs
  - *Identify responsive hosts*
- Second phase: Application layer
  - OS-level TCP handshake
  - *Deliver payloads & reconnaissance*



Two-phase Scanning

- First phase: Transport layer
  - Hand-crafted, stateless SYNs
  - *Identify responsive hosts*
- Second phase: Application layer
  - OS-level TCP handshake
  - *Deliver payloads & reconnaissance*



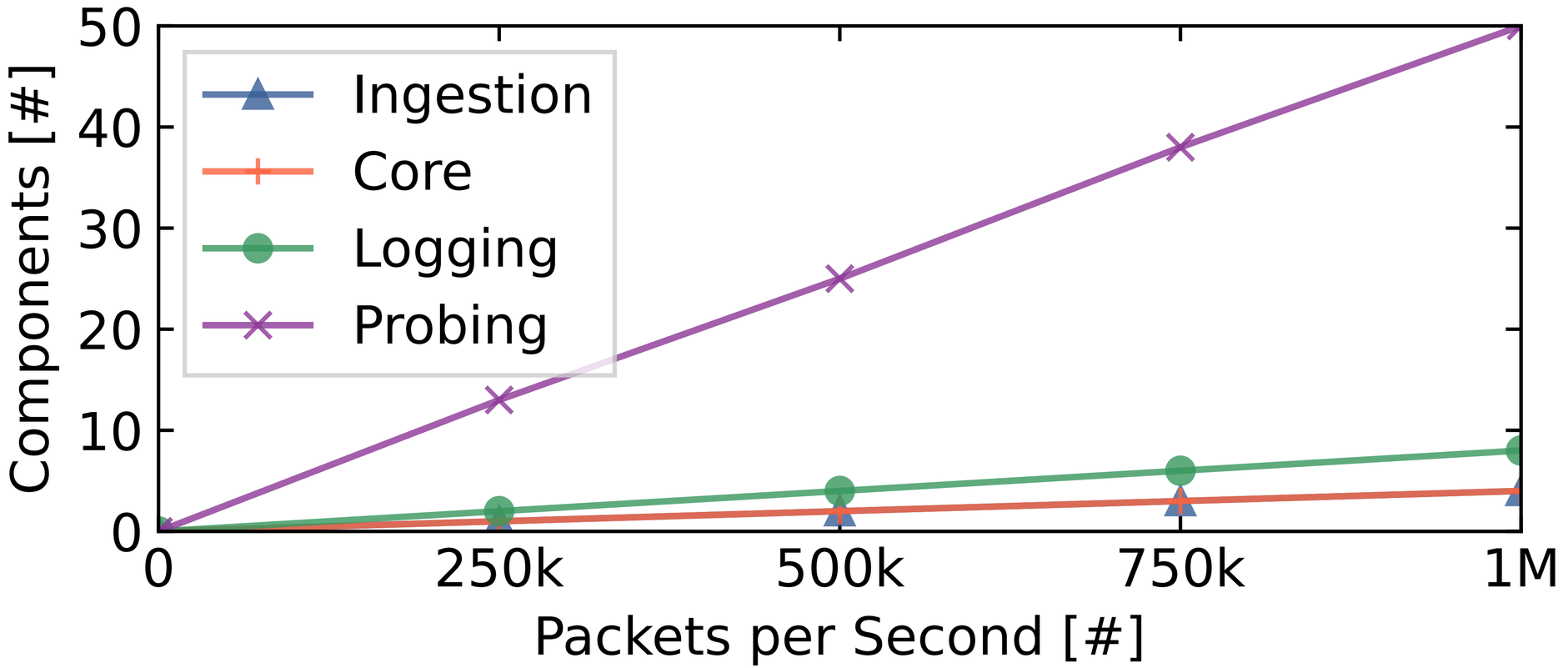
Spoki: Revealing Two-phase Scanners

- Spoki interacts with two-phase scanners in real time
  - Scalable system based on actors with the C++ Actor Framework (CAF)
  - Libtrace for packet ingestion, Scamper for probing
  - Collects payloads after accepting TCP connections
- Deployed in two /24 prefixes (US, EU)

- Published source code on GitHub (https://github.com/inetrg/spoki)

Spoki rate-limits probes and uses small packets to avoid participating in DoS.

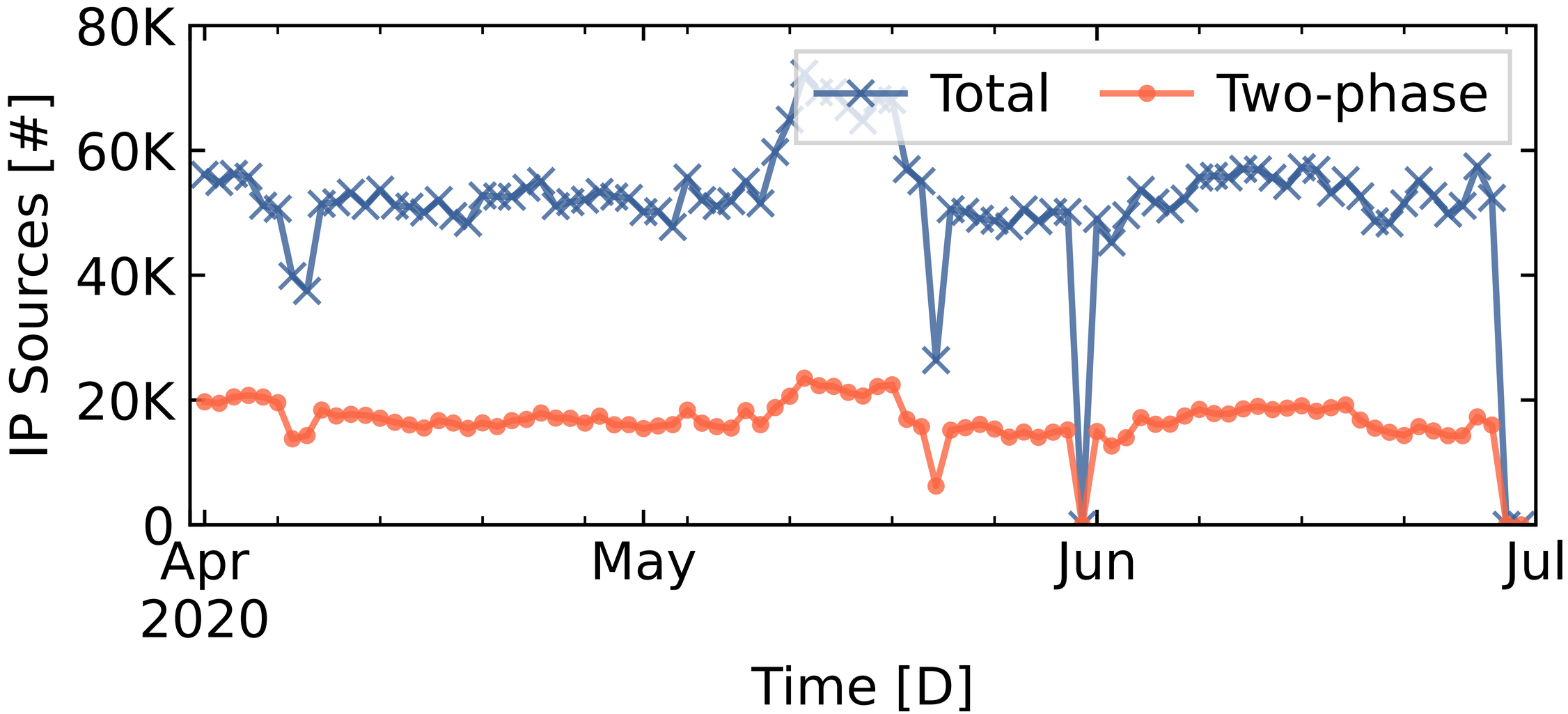
Scaling Up to 1 Million Probes Per Second



Parallel components allow Spoki to process large traffic volumes.

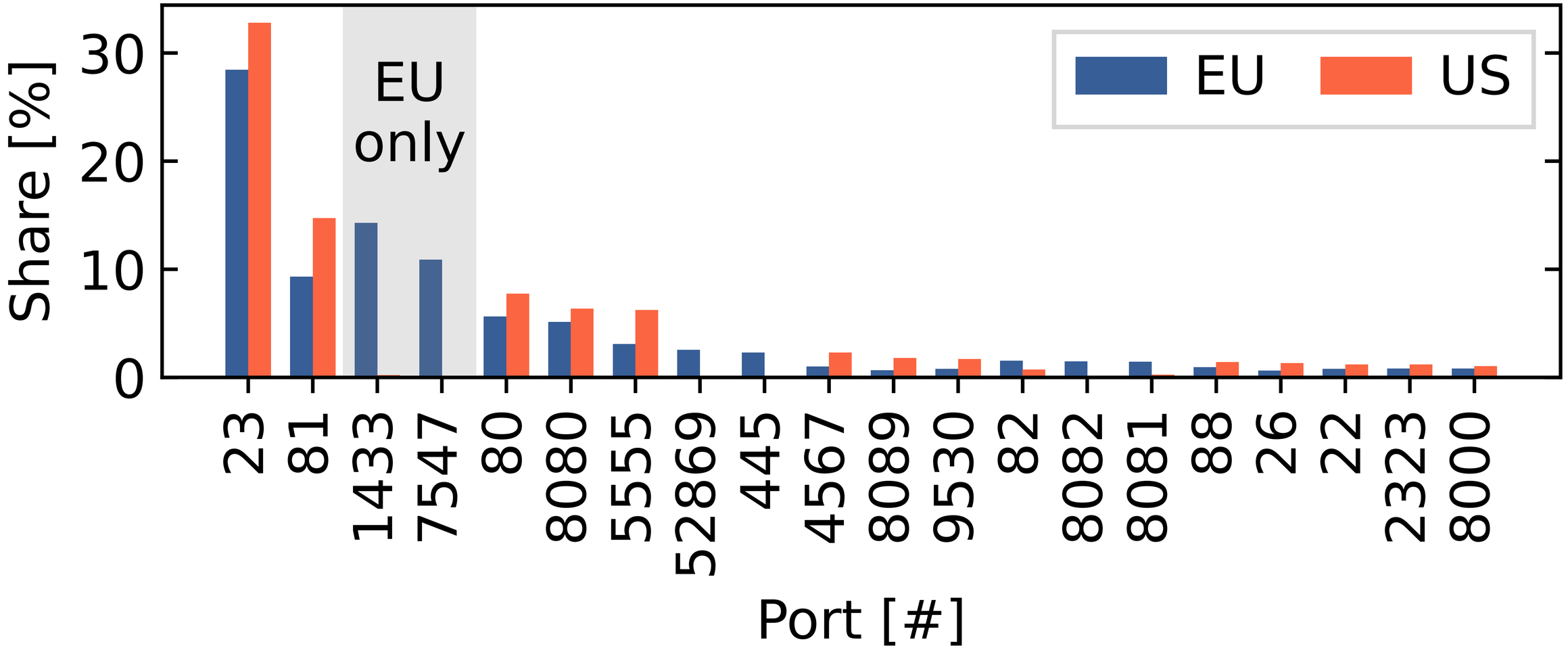
Share of Two-phase Sources

About 30% of sources send two-phase events each day.



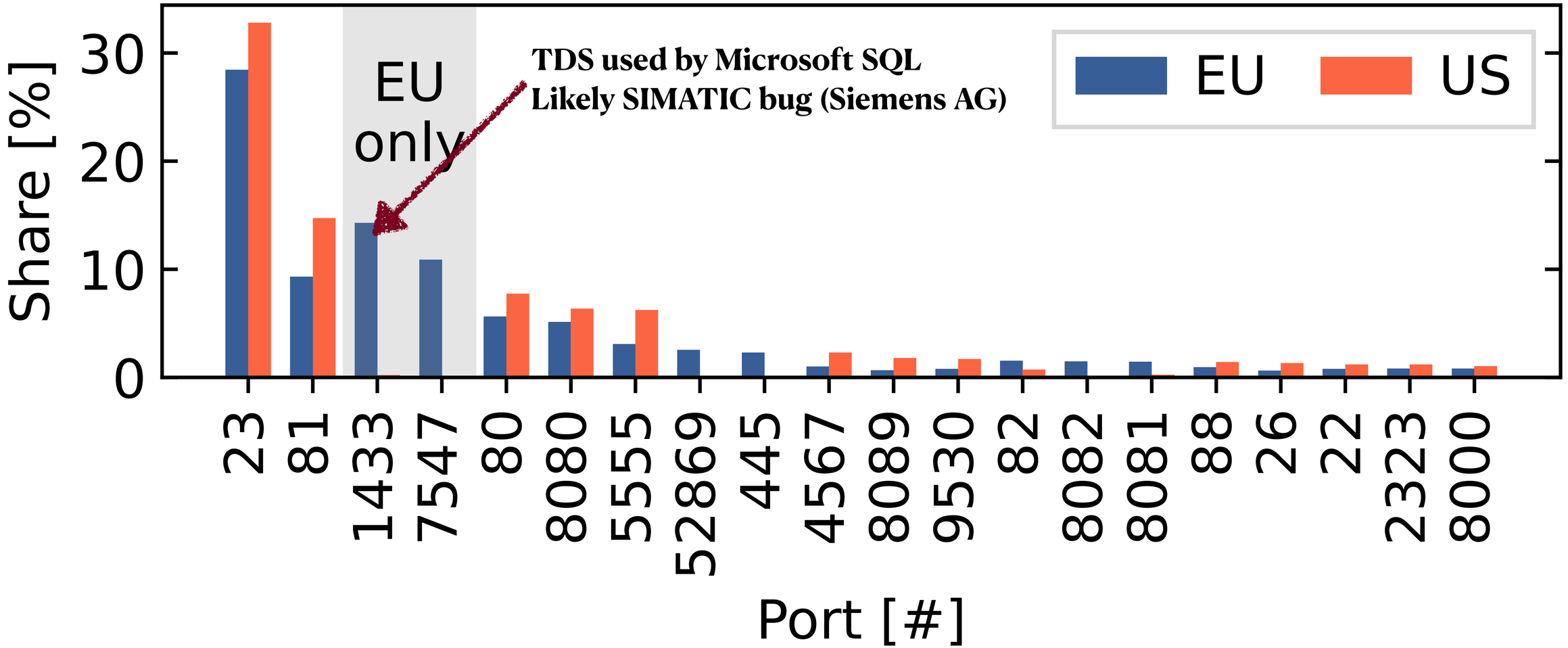
Targeted Ports

Two ports are scanned exclusively in the EU.



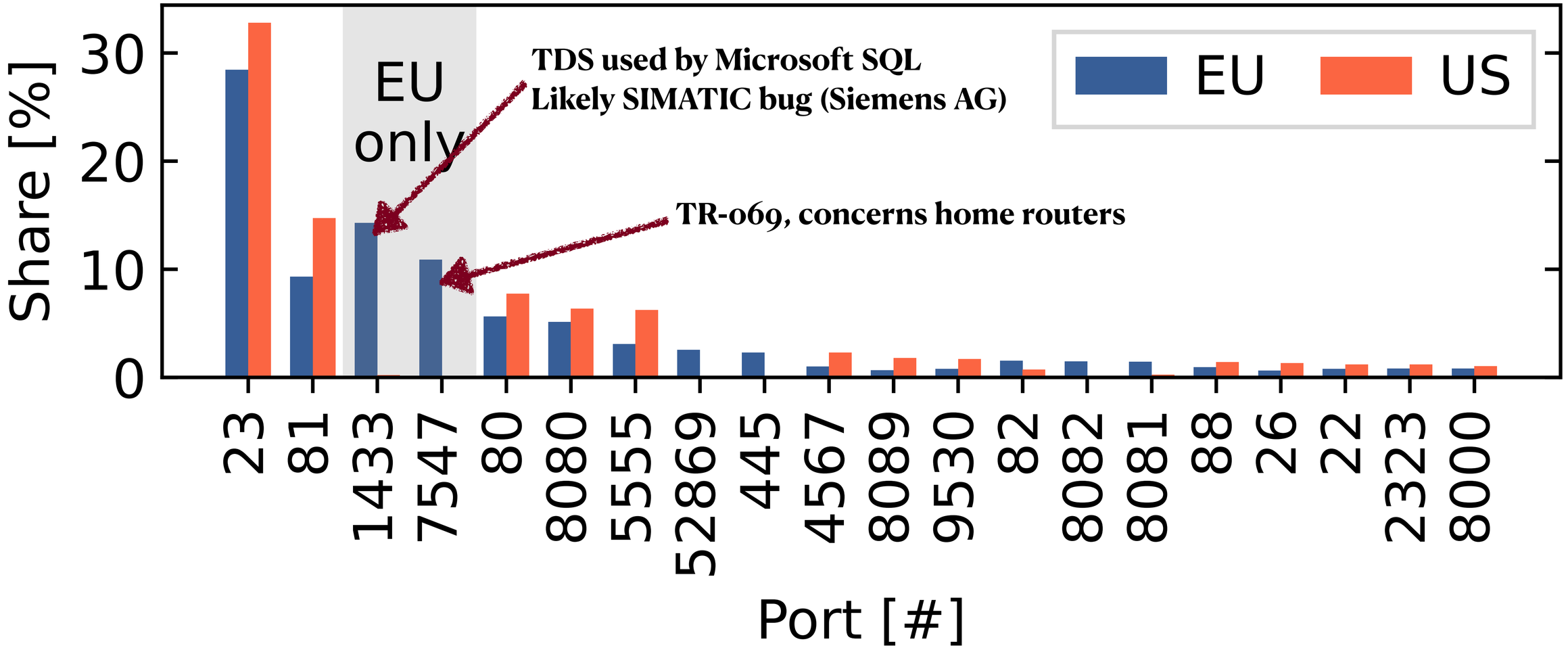
Targeted Ports

Two ports are scanned exclusively in the EU.



Targeted Ports

Two ports are scanned exclusively in the EU.



TCP Payloads

- TCP payloads are not available in traditional telescopes
- We scan payloads for *downloaders*: shell code that downloads malware

Event Type	EU		US	
ASCII	2,155,751	58.6%	1,984,444	80.4%
HEX	1,478,556	40.2%	339,217	13.8%
Downloader	42,303	1.2%	143,309	5.8%

- Sample names and types match known malware such as the Mozi P2P-botnet
- Spoki detected 15% of the samples earlier than VirusTotal (26% benign, 59% old)

The Maliciousness of Two-Phase Scanners

Malware distribution clearly points at malicious intent. Can we validate our findings?

The Maliciousness of Two-Phase Scanners

Malware distribution clearly points at malicious intent. Can we validate our findings?

Approach 1: Semi-Manual Analysis

- Reveals malicious payloads such as:

Port	Attack
1433	TDS, SQL, SIMATIC
7545	TR-069, routers
5555	ADB crypto miner
9530, 4567	Embedded devices
5432	Realtek UPnP

The Maliciousness of Two-Phase Scanners

Malware distribution clearly points at malicious intent. Can we validate our findings?

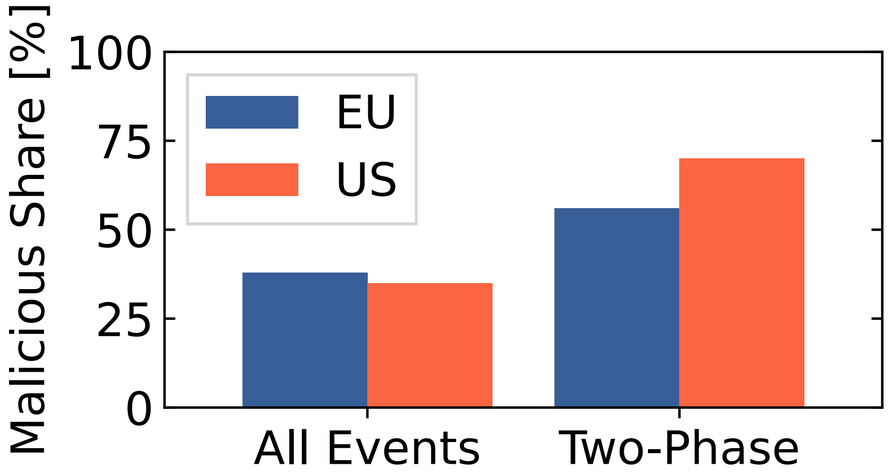
Approach 1: Semi-Manual Analysis

- Reveals malicious payloads such as:

Port	Attack
1433	TDS, SQL, SIMATIC
7545	TR-069, routers
5555	ADB crypto miner
9530, 4567	Embedded devices
5432	Realtek UPnP

Approach 2: Query GreyNoise

- Classifies IPs into: *malicious*, *benign*, and *unknown*
- Two-phase events have a higher share of malicious sources:



Geographical Scanning Locality

- Scanners focus on different ports in Europe and the USA
- Different vendors and deployments attracts different attacks

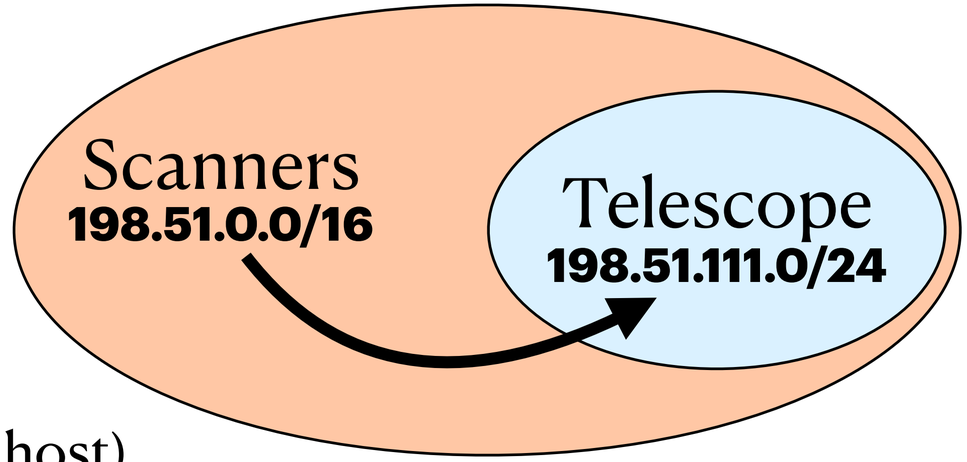
	EU		US	
Payload Prefix	**Share**	**Ports**	**Share**	**Ports**
TDS7 Pre-login	74.52%	1433	1.16%	1443
TLS Client Hello	4.55%	443, 8443	37.80%	443, 8443
ADB Connect	4.97%	5555	37.01%	5555
SMB Negotiate	11.04%	445	–	
PSQL/UPnP	0.35%	5432	3.10%	5432, 5000
TSAP	0.45%	102	1.42%	102
MongoDB	0.27%	27017	1.21%	27017
Unknown	0.16%	28967	1.15%	28967

Targets non-ASCII payloads

TDS: Tabular Data Stream used by Microsoft SQL
ADB: Android Debug Bridge
TSAP: Transport Service Access Point protocol port, used for x.400, X.500; vulnerabilities in a variety of SIMATIC devices

Topological Scanning Locality

- Six of the top-ten source prefixes in the EU share a /16 with our /24 vantage point
  - This scanning behavior is associated with botnets
  - A similar locality cannot be observed in the US



- Crosscheck (sampled) traffic at a European IXP
  - Local, irregular SYNs in 370 prefixes (150 packets per host)
  - Very focused: 96% target 23, 7547, 8291 (multiple sources identified as MiktoTik routers)

- No correlation of /16 local, irregular SYNs at an Asian ISP

Takeaways

- Spoki makes two-phase scanners visible
- Irregular SYNs dominate SYNs on the Internet: ~75%
- Two-phase scans
  - … act as a catalyst
  - … are used for malicious activities
  - … follow locality patterns
  - … have detectable signatures

Takeaways

- Spoki makes two-phase scanners visible
- Irregular SYNs dominate SYNs on the Internet: ~75%
- Two-phase scans
  - … act as a catalyst
  - … are used for malicious activities
  - … follow locality patterns
  - … have detectable signatures

➡ Short update cycles needed

Takeaways

- Spoki makes two-phase scanners visible
- Irregular SYNs dominate SYNs on the Internet: ~75%
- Two-phase scans
  - … act as a catalyst → Short update cycles needed
  - … are used for malicious activities → Deliver a variety of malware
  - … follow locality patterns
  - … have detectable signatures

Takeaways

- Spoki makes two-phase scanners visible
- Irregular SYNs dominate SYNs on the Internet: ~75%
- Two-phase scans
  - … act as a catalyst → Short update cycles needed
  - … are used for malicious activities → Deliver a variety of malware
  - … follow locality patterns → Ensure your data fits your deployment
  - … have detectable signatures

Takeaways

- Spoki makes two-phase scanners visible
- Irregular SYNs dominate SYNs on the Internet: ~75%
- Two-phase scans
  - … act as a catalyst ➞ Short update cycles needed
  - … are used for malicious activities ➞ Deliver a variety of malware
  - … follow locality patterns ➞ Ensure your data fits your deployment
  - … have detectable signatures ➞ Can be tracked and their packets filtered

Thank you for your attention!

Find the paper, code, and artifacts at:
https://spoki.secnow.net

Contact: raphael.hiesgen@haw-hamburg.de