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

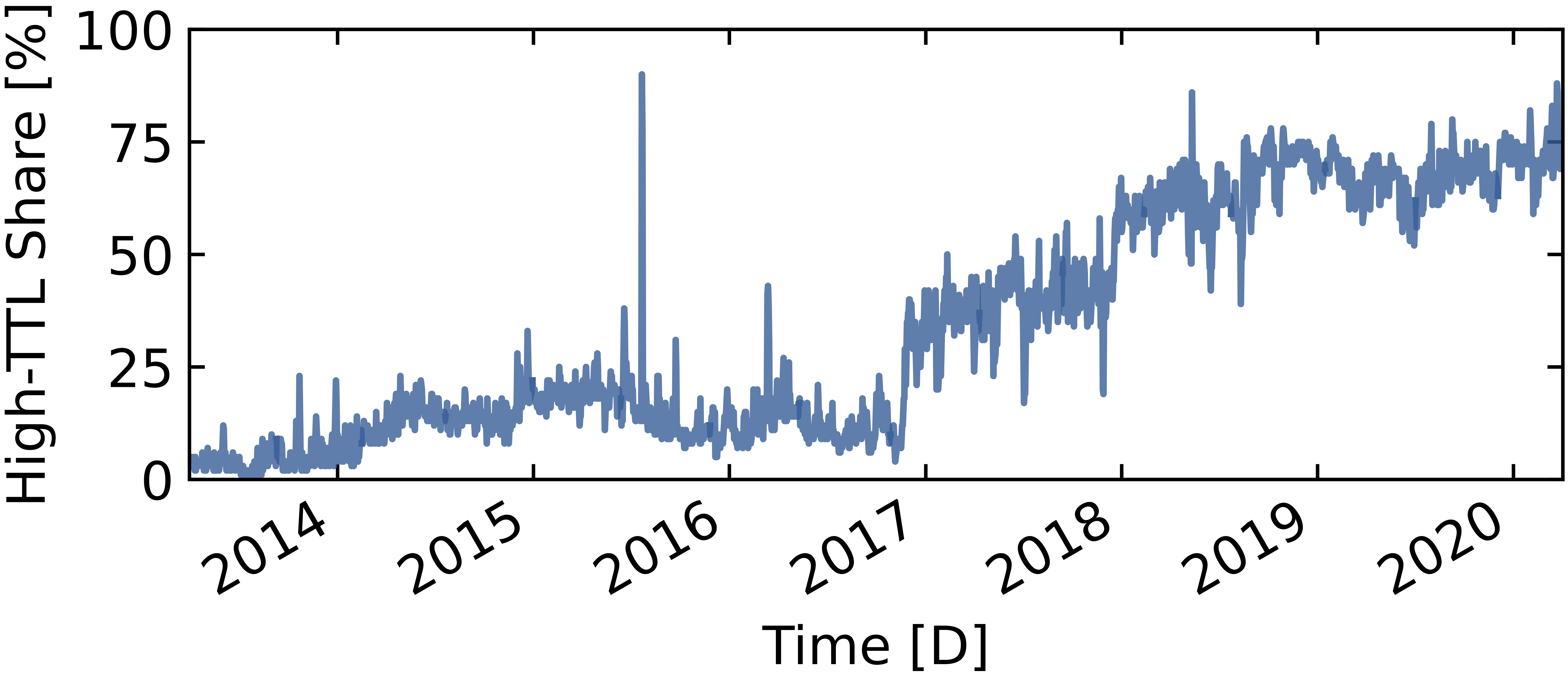
HAW HAMBURG

Freie Universität Berlin

caida

The Share of Irregular Packets is Increasing

UCSD Network Telescope



Agenda

Two-phase Scanners

Spoki

Behavior

Payloads

Locality

What is a TCP Irregularity?

- Irregular packets show one or more of:
  - High TTL (≥200)
  - No TCP options
  - Striking IP ID (54321)
- The telescope now observes a share of roughly 75% irregular SYNs

What is a TCP Irregularity?

- Irregular packets show one or more of:
  -
  -
  -
- T

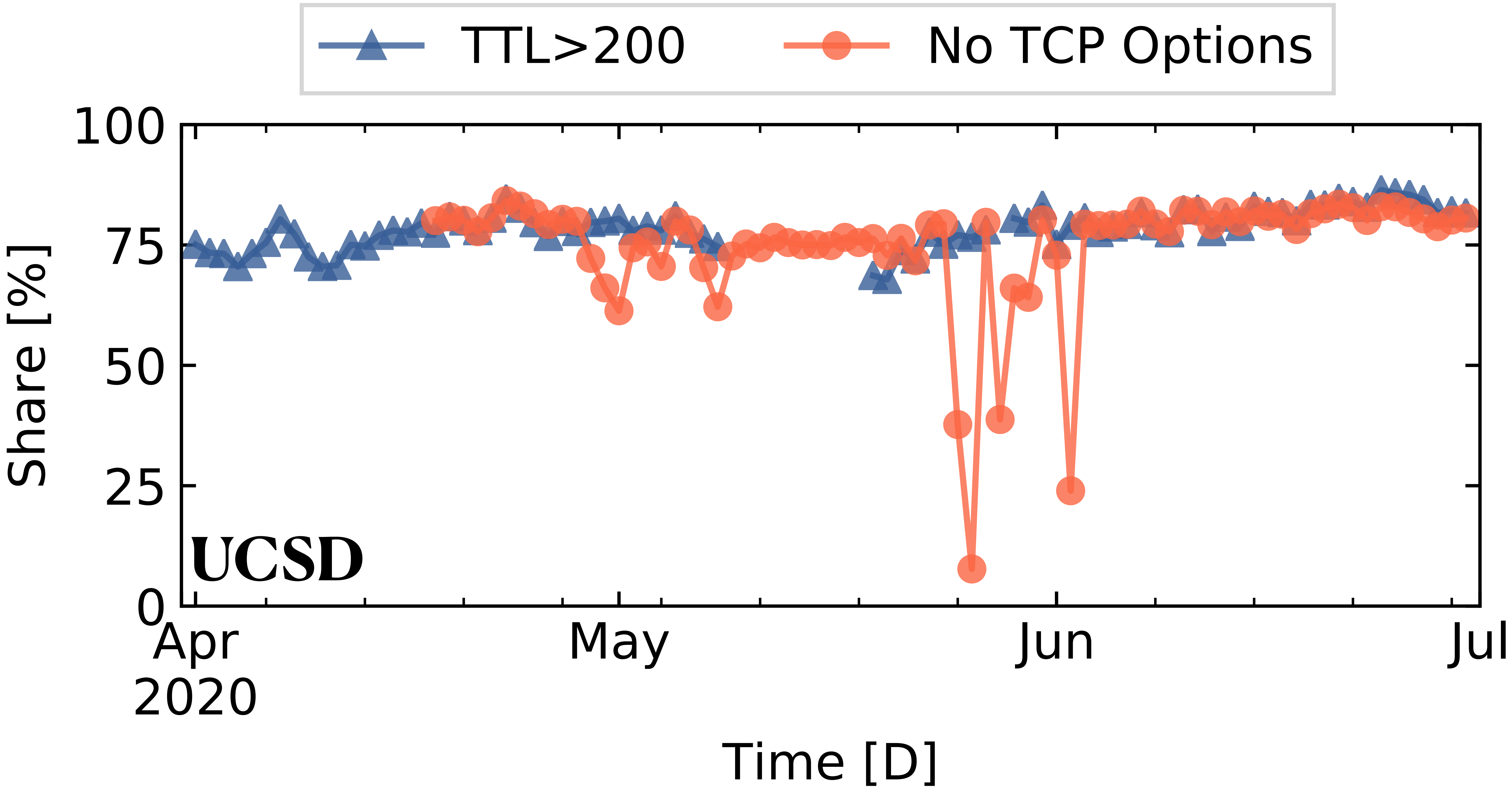
Is this observation specific to the UCSD network telescope?

A Global Phenomenon

- We observe this at three vantage points
- TTL and TCP opts. share largely overlap

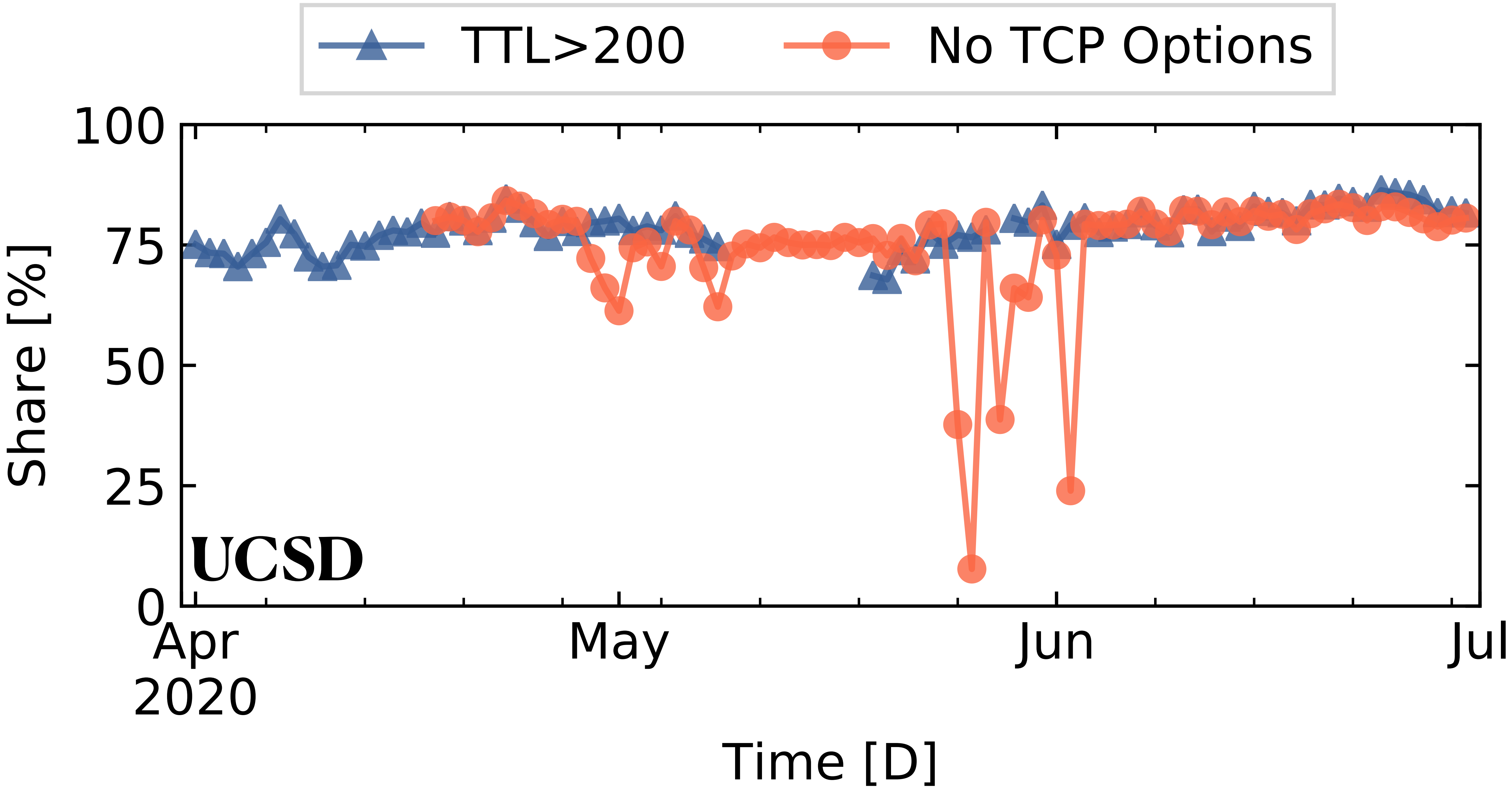
A Global Phenomenon

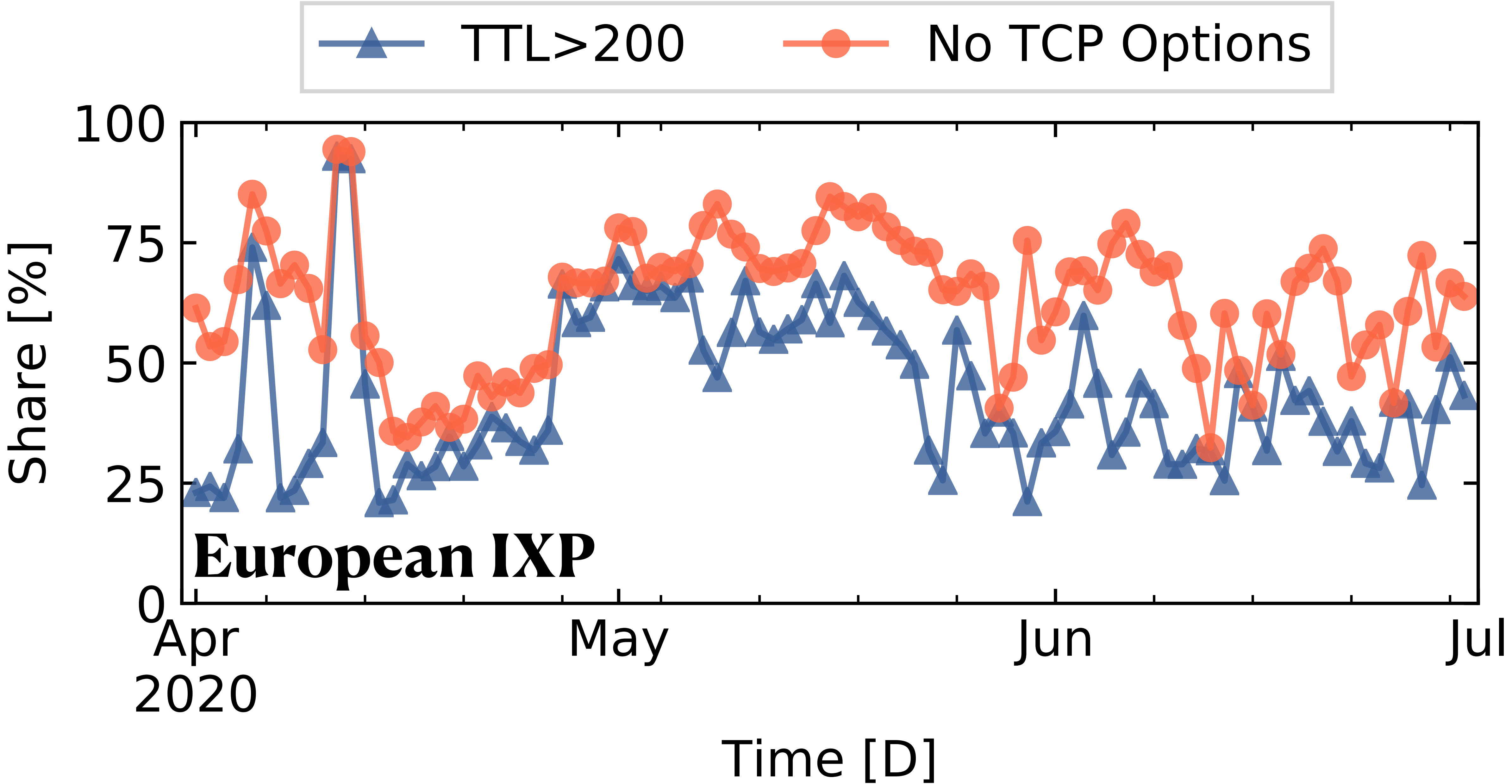
- We observe this at three vantage points
- TTL and TCP opts. share largely overlap



A Global Phenomenon

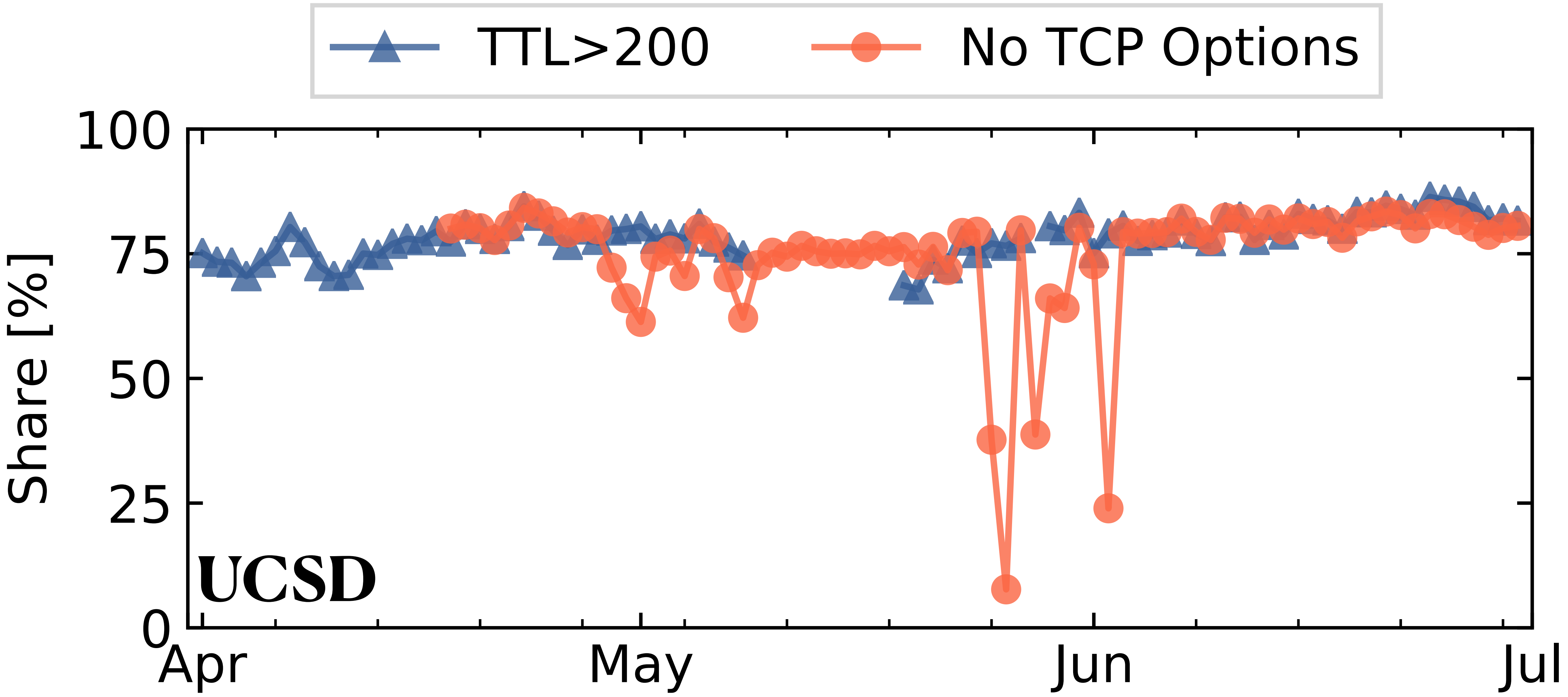
- We observe this at three vantage points
- TTL and TCP opts. share largely overlap

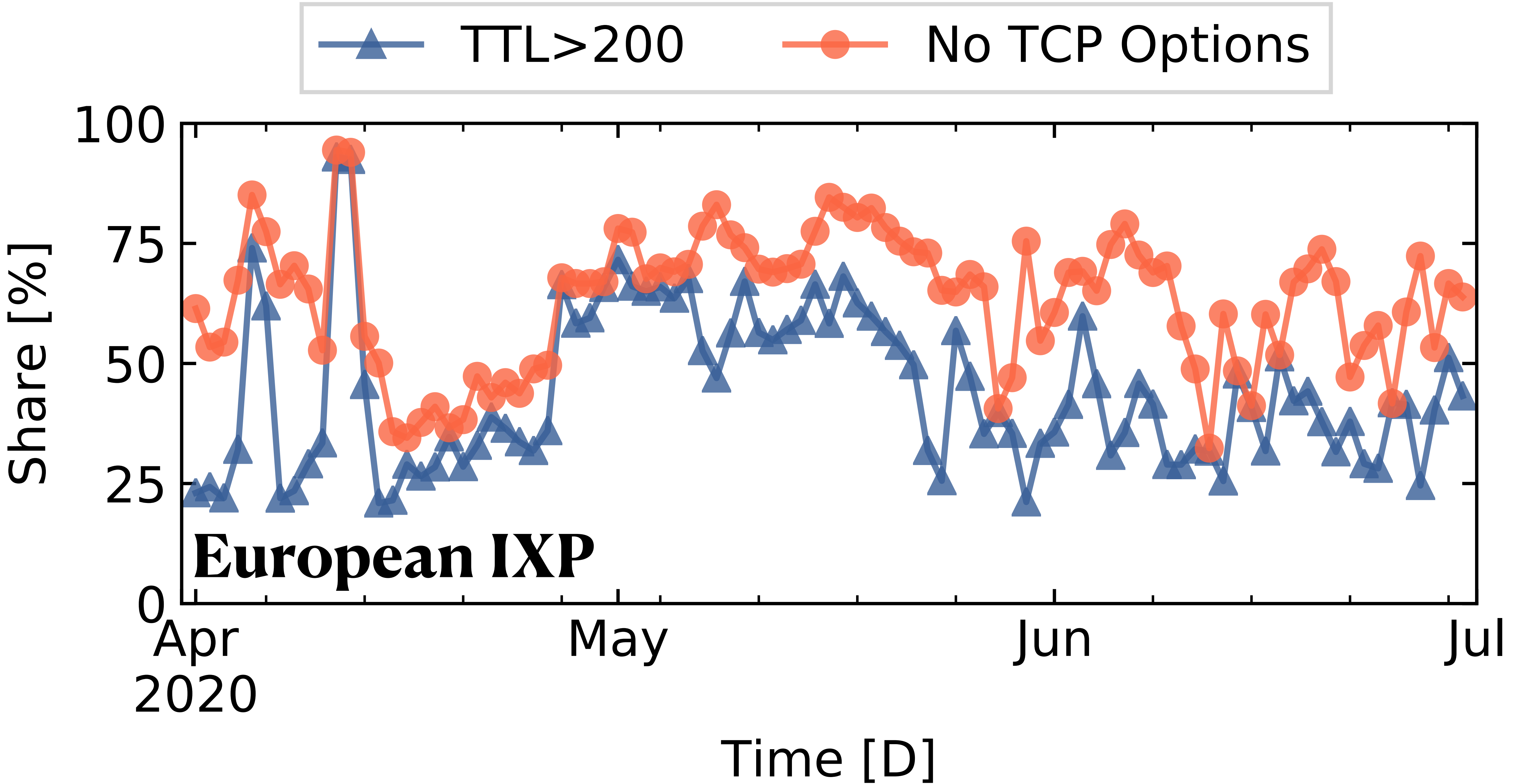


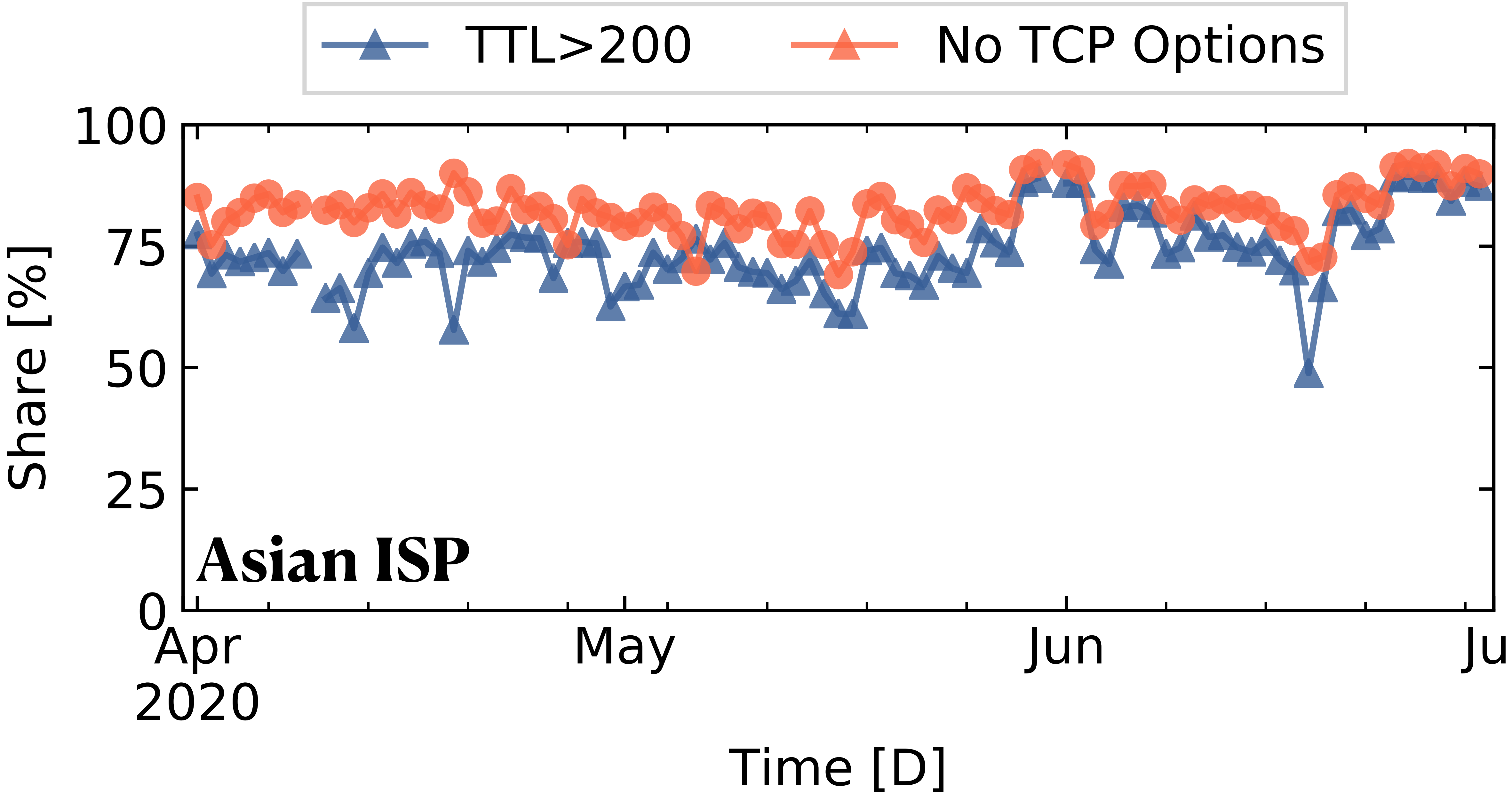


A Global Phenomenon

- We observe this at three vantage points
- TTL and TCP opts. share largely overlap

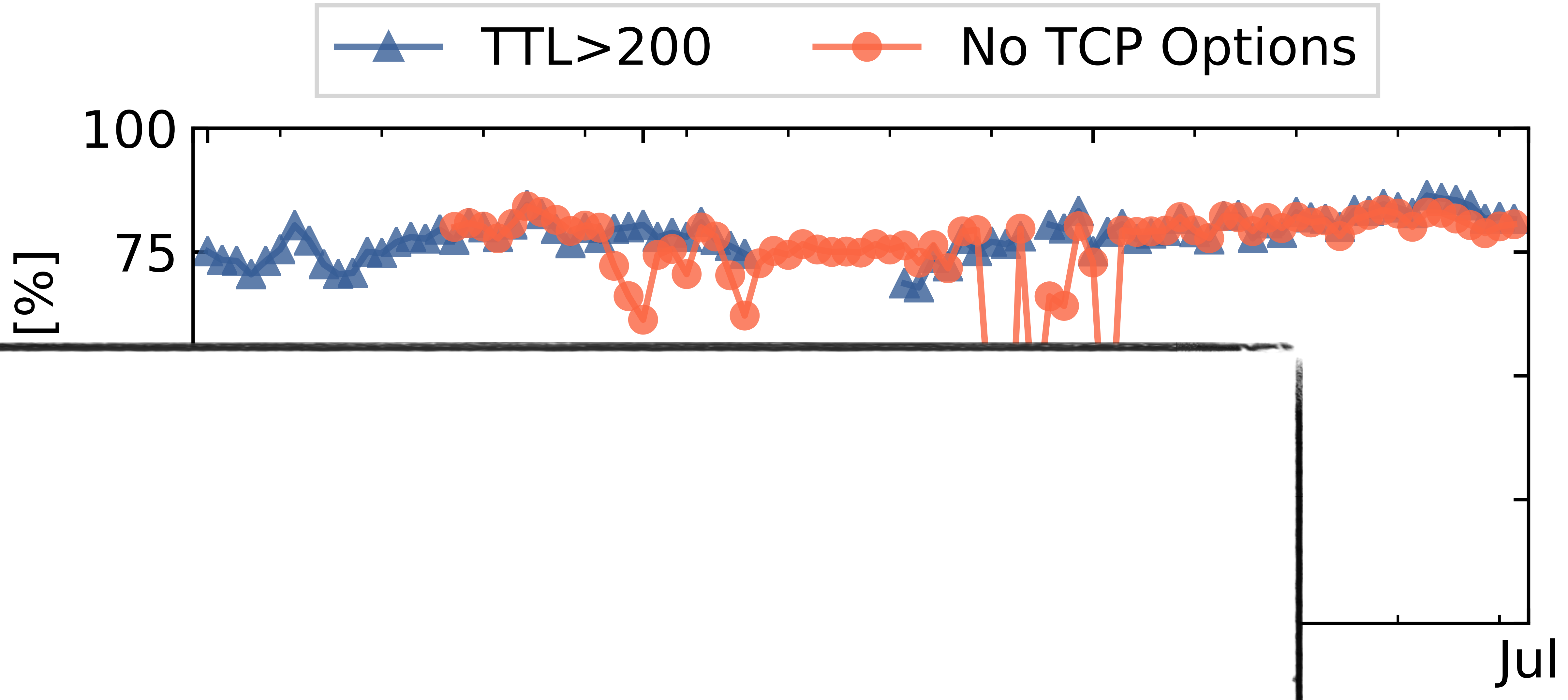




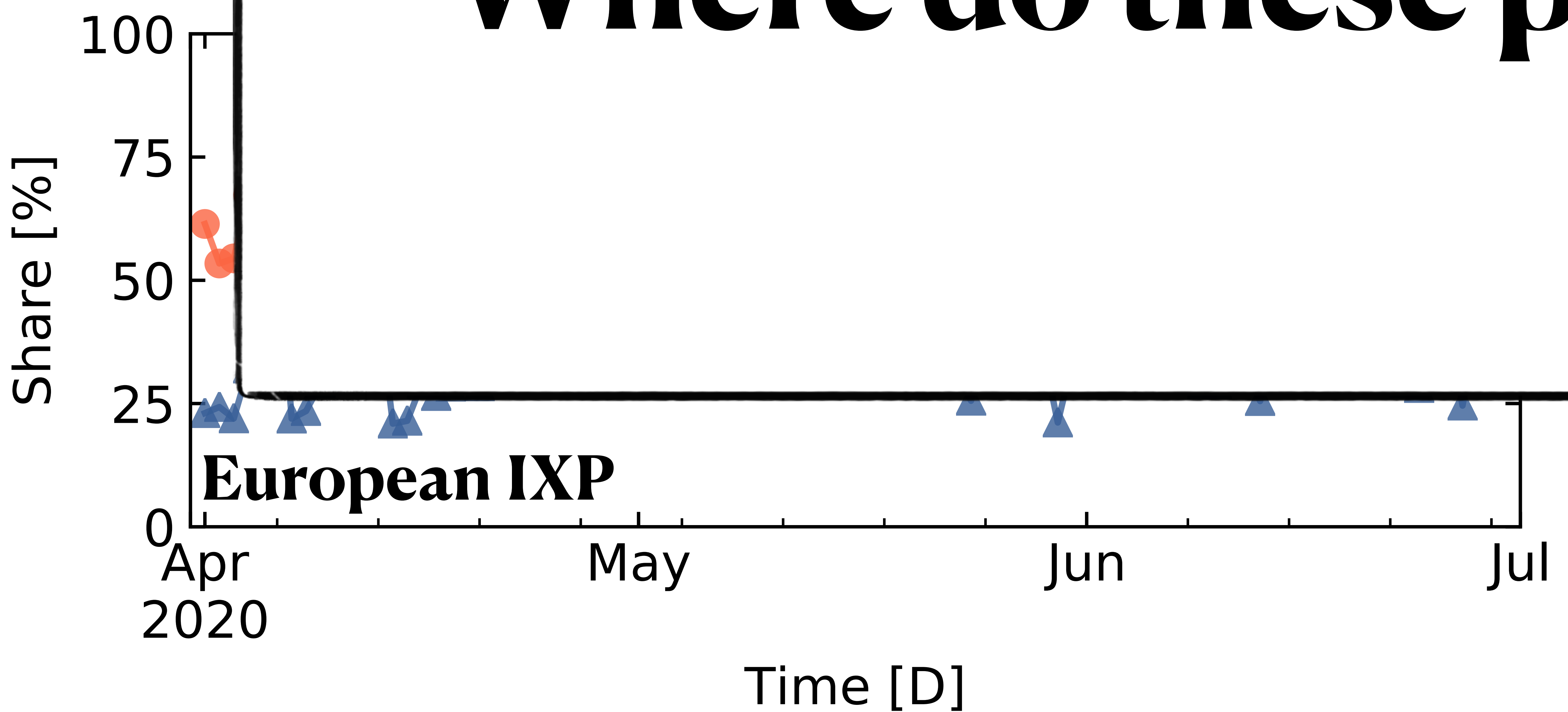


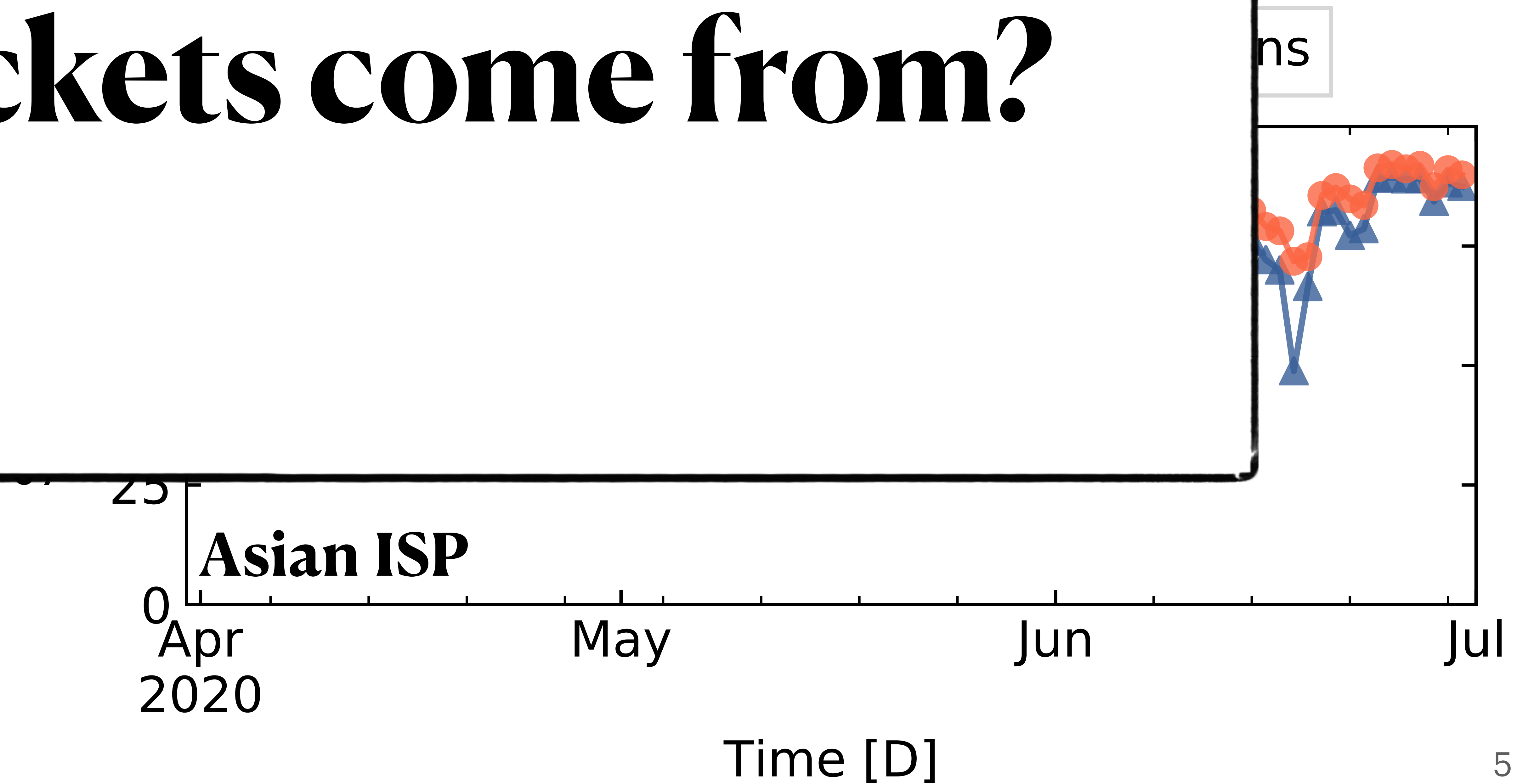
A Global Phenomenon

- We observe this at three vantage points
- T



Where do these packets come from?





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

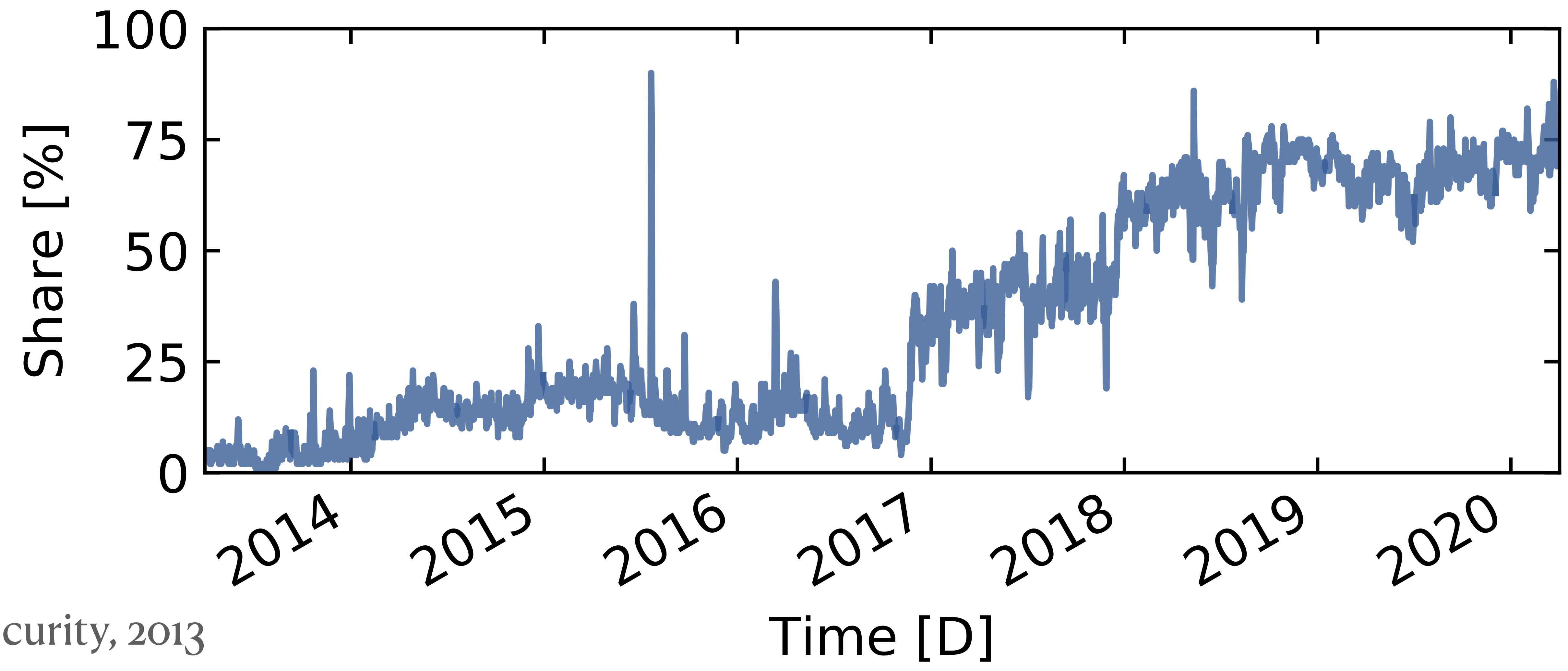
Durumeric et al., *ZMap: Fast Internet-Wide Scanning and its Security Applications,* USENIX Security, 2013
Antonakakis et al., *Understanding the Mirai Botnet,* USENIX Security, 2017

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



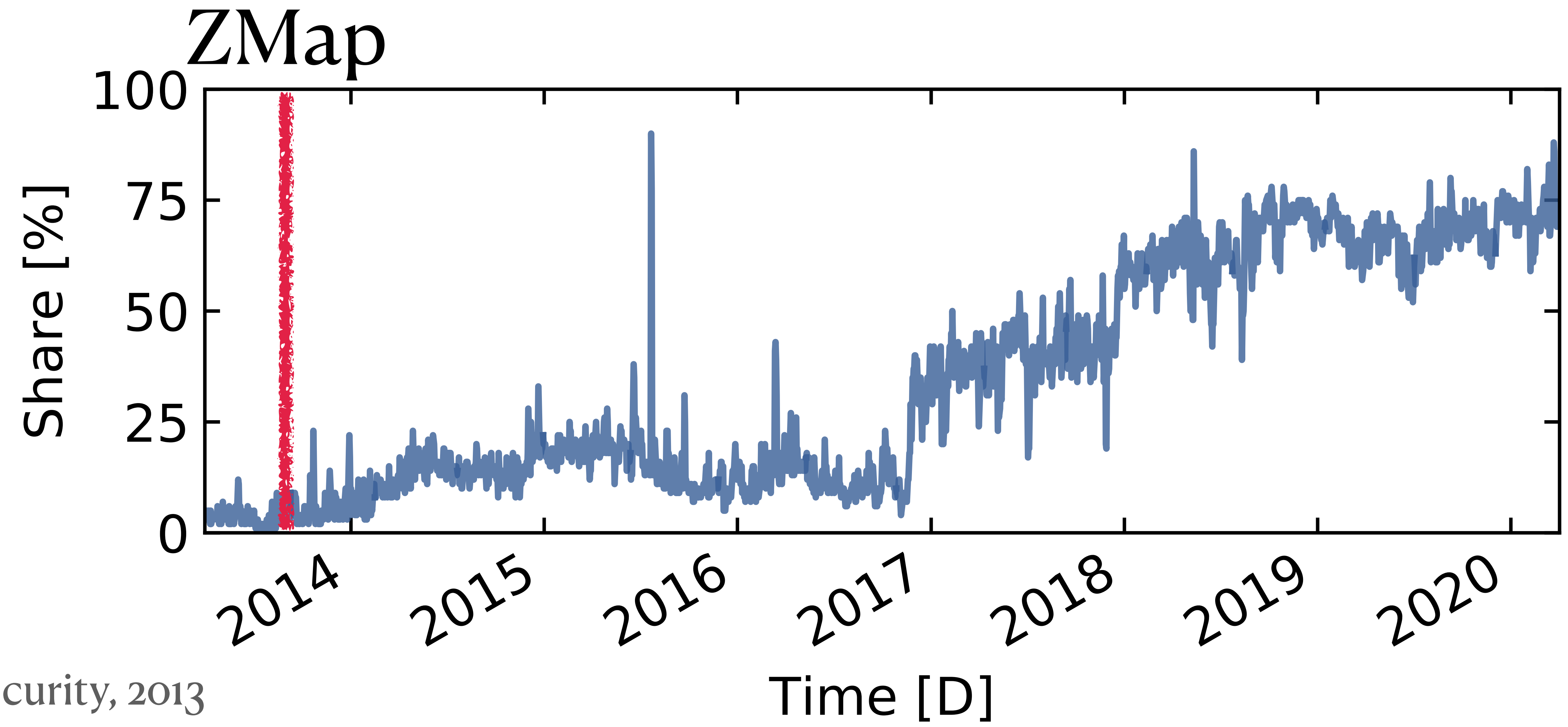
Durumeric et al., *ZMap: Fast Internet-Wide Scanning and its Security Applications*, USENIX Security, 2013
Antonakakis et al., *Understanding the Mirai Botnet*, USENIX Security, 2017

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



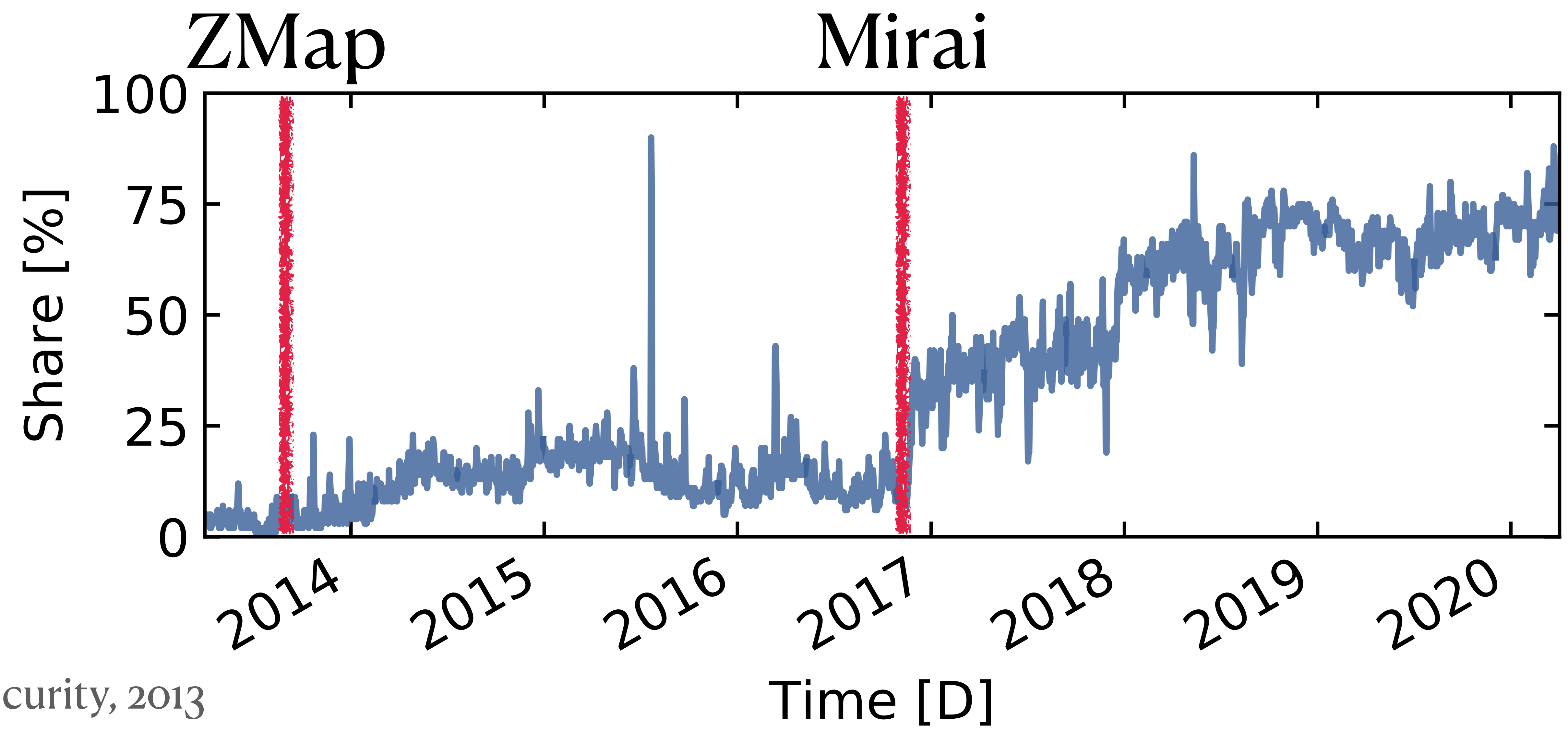
Durumeric et al., *ZMap: Fast Internet-Wide Scanning and its Security Applications*, USENIX Security, 2013
Antonakakis et al., *Understanding the Mirai Botnet*, USENIX Security, 2017

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



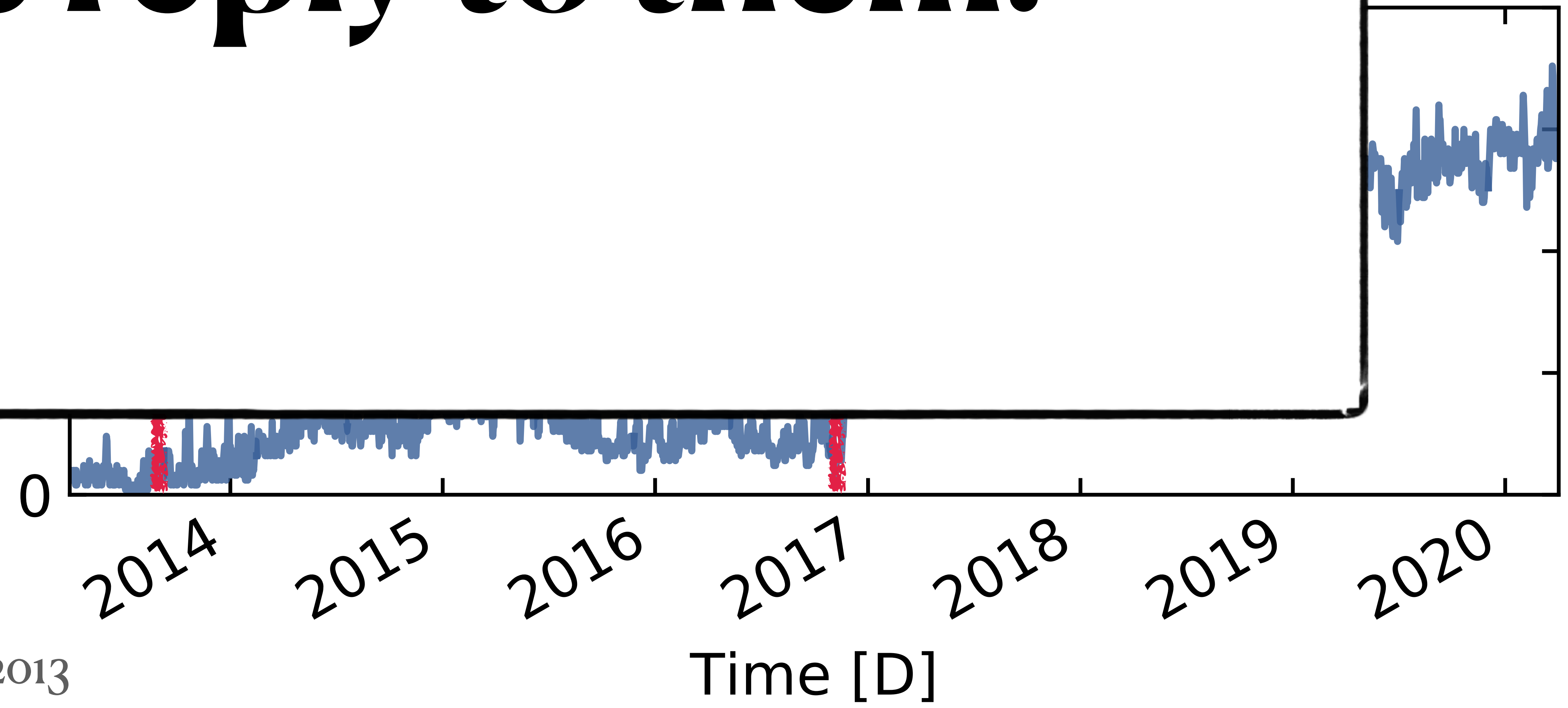
Durumeric et al., *ZMap: Fast Internet-Wide Scanning and its Security Applications*, USENIX Security, 2013
Antonakakis et al., *Understanding the Mirai Botnet*, USENIX Security, 2017

Background: Stateless Scanning

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

- Increases scan speeds by avoiding local state
  -
  -
- P
- A

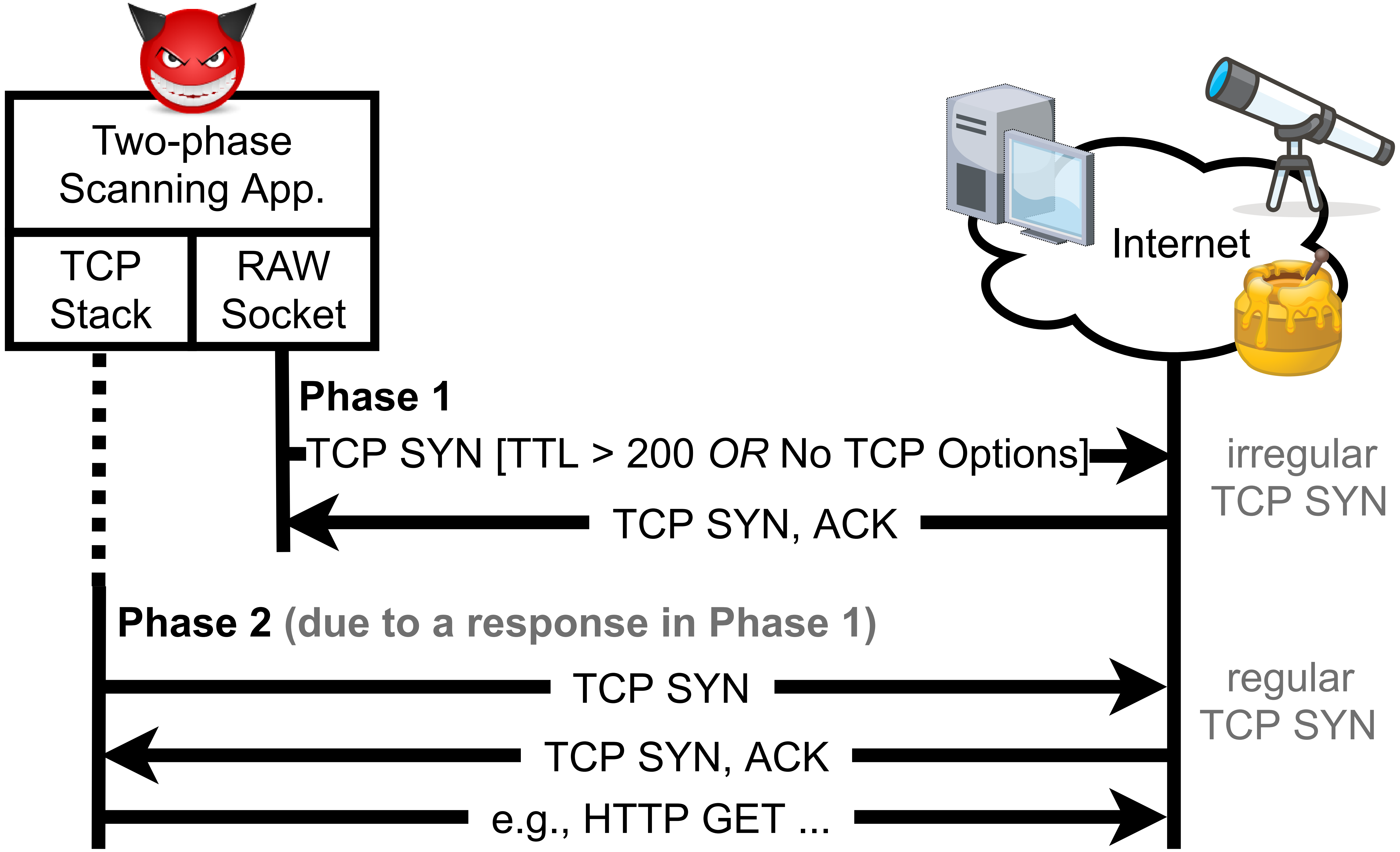
What happens if we reply to them?



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

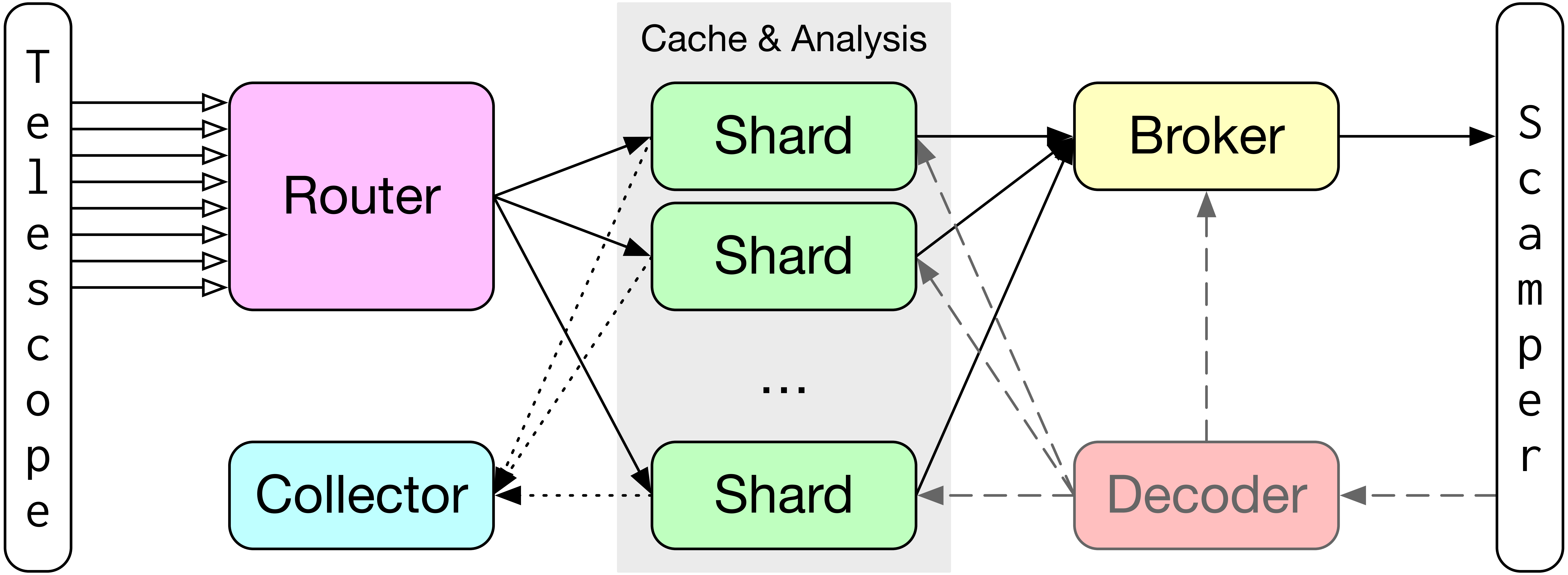


Spoki: Revealing Two-phase Scanners

- 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

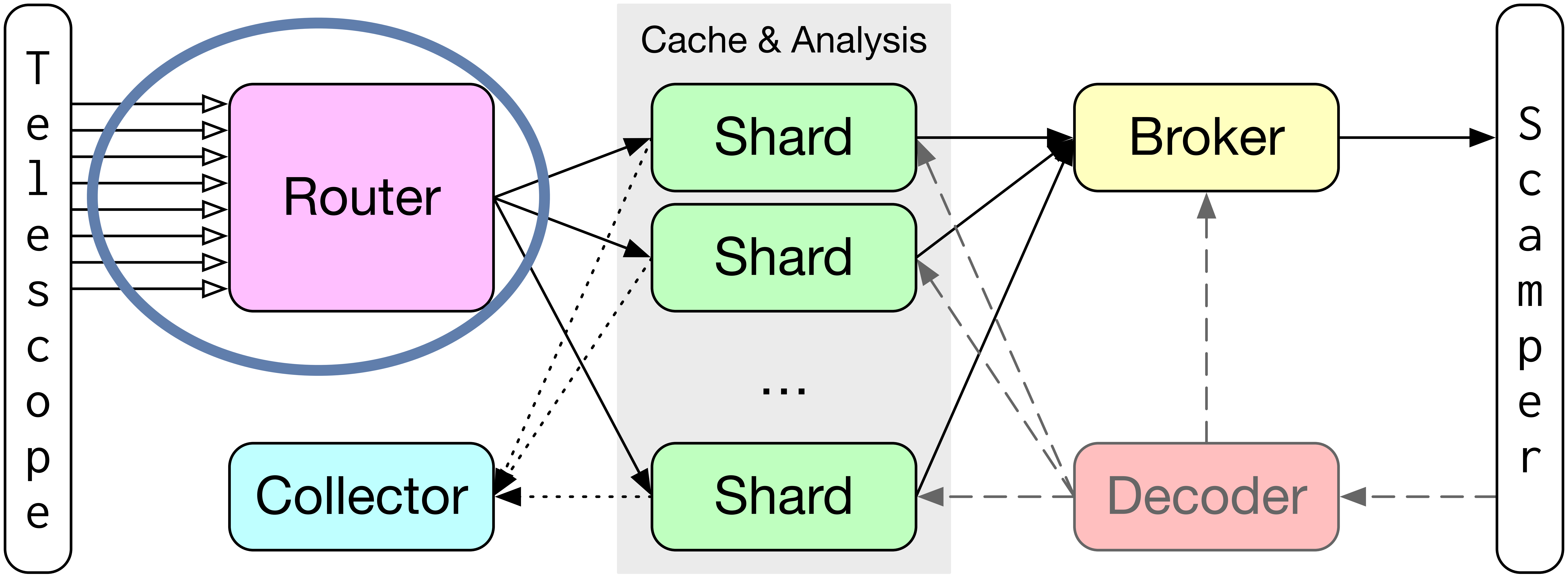
Spoki rate-limits probes and uses small packets to avoid participating in DoS.

Architecture of Spoki

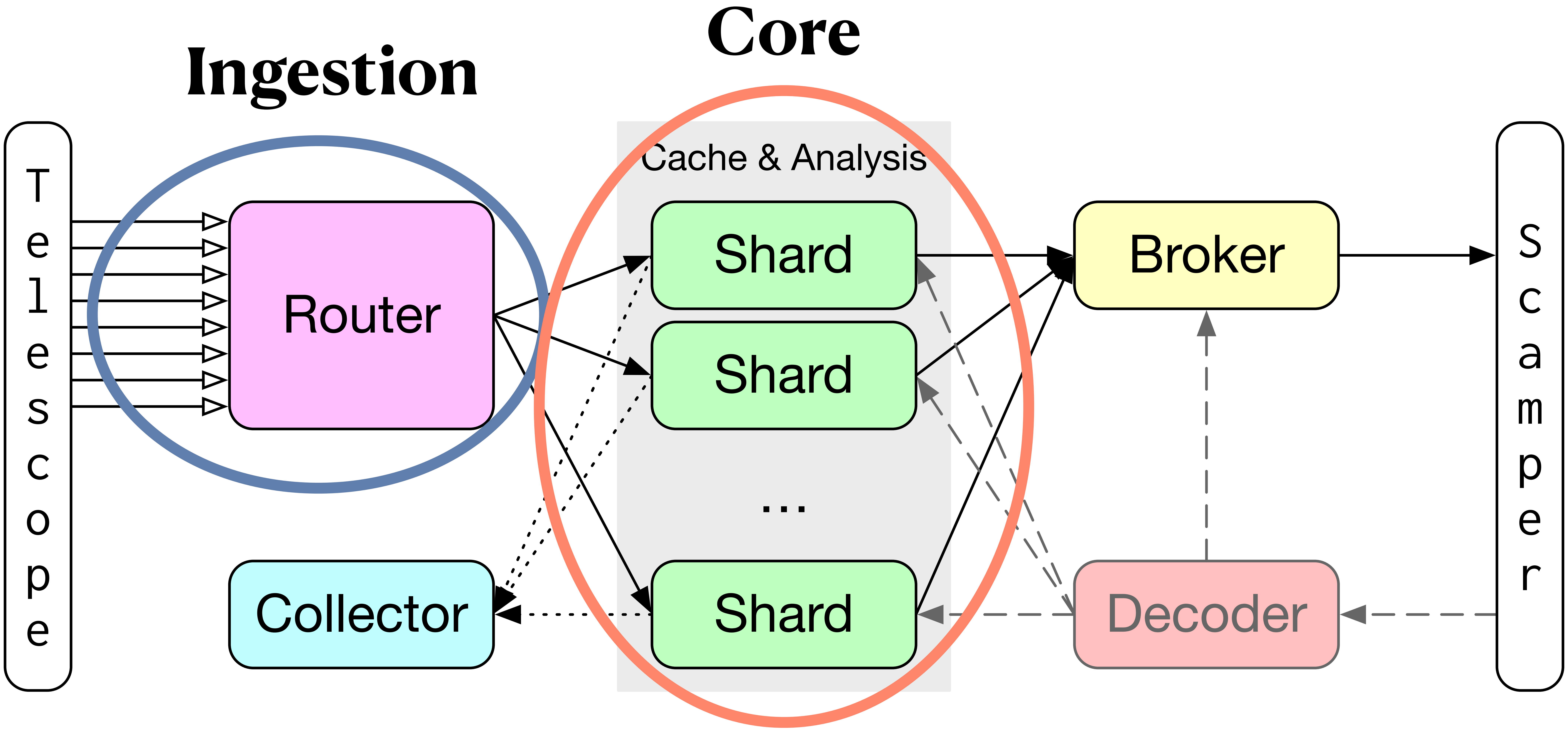


Architecture of Spoki

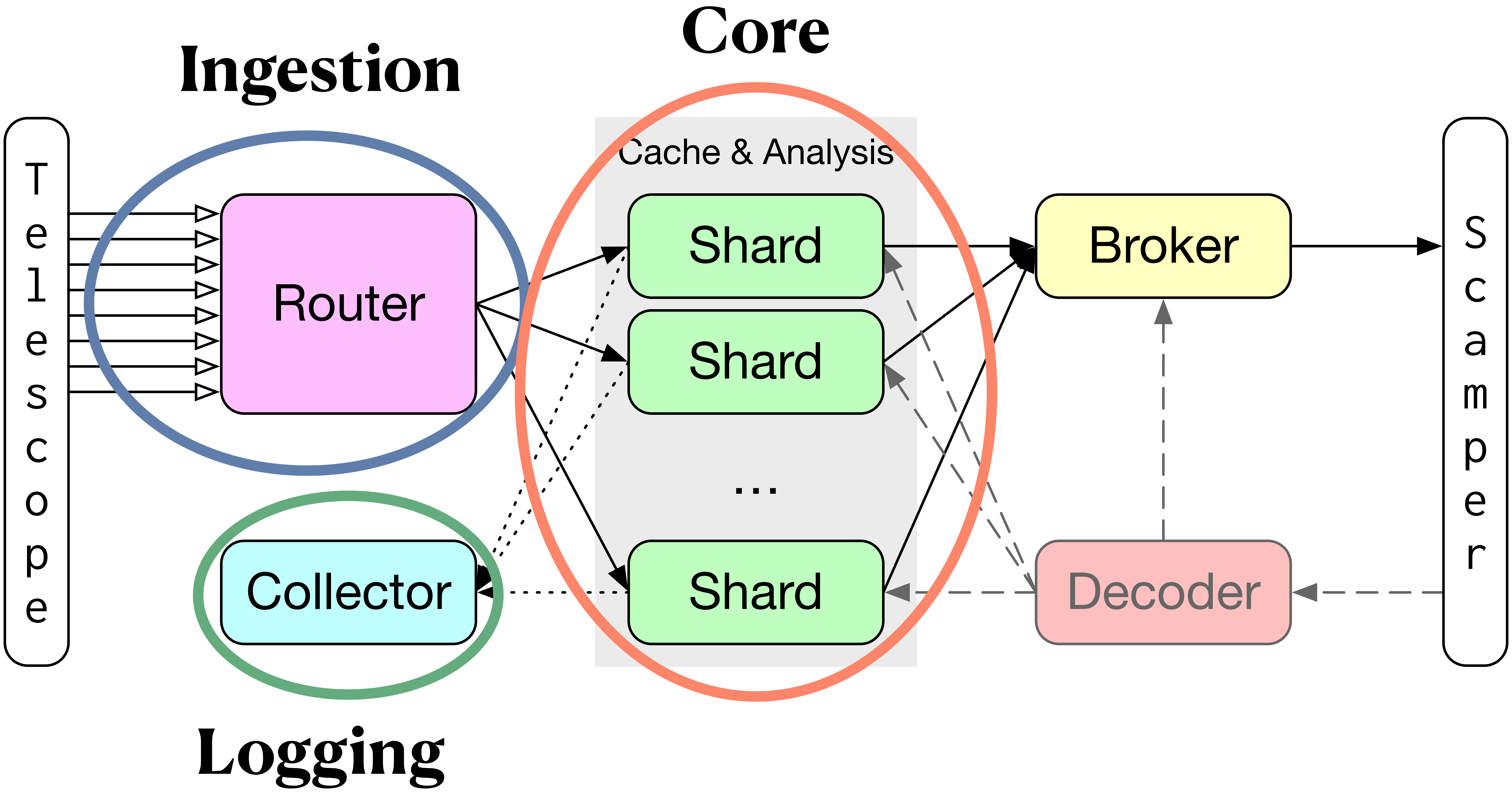
Ingestion



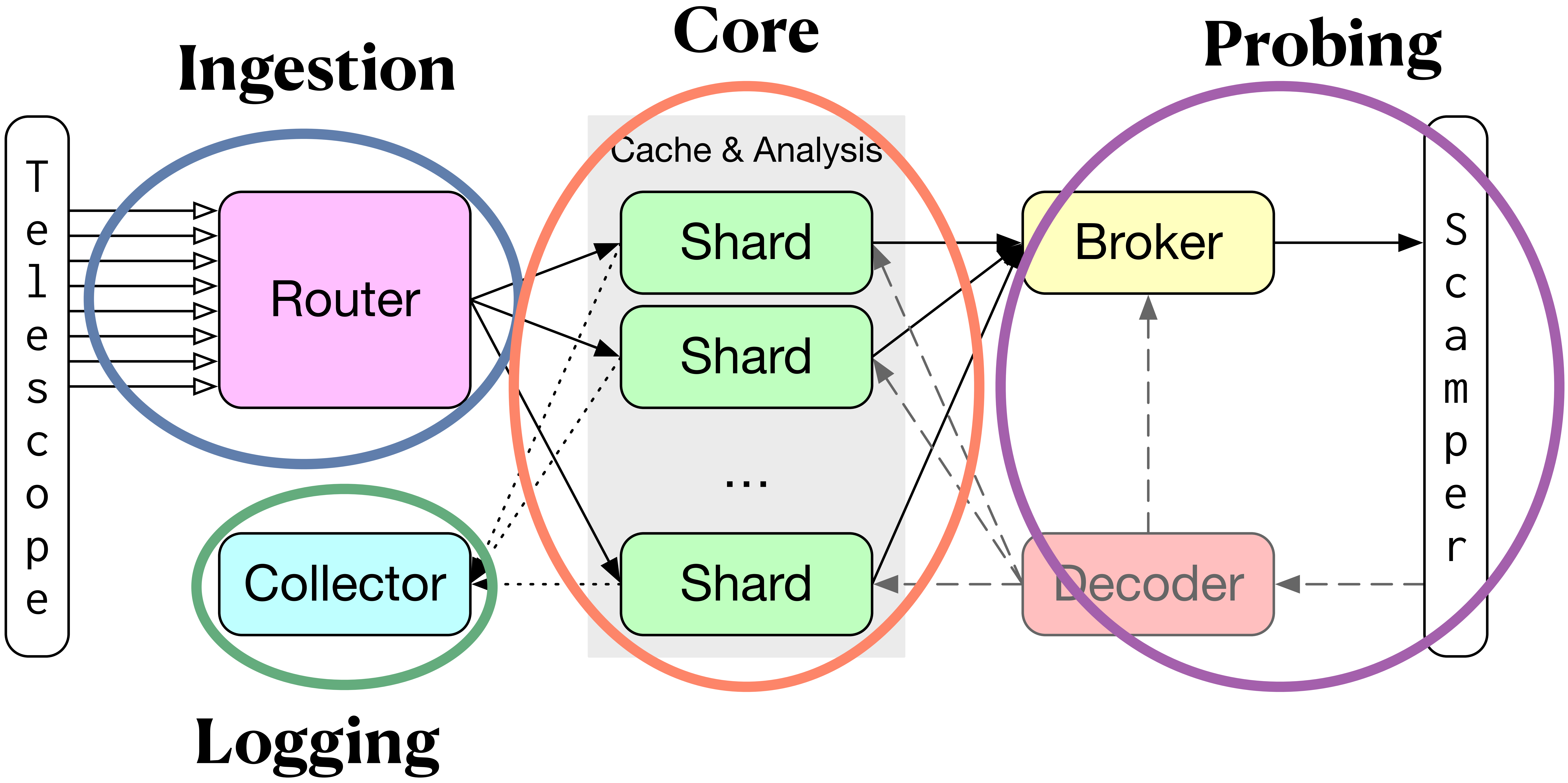
Architecture of Spoki



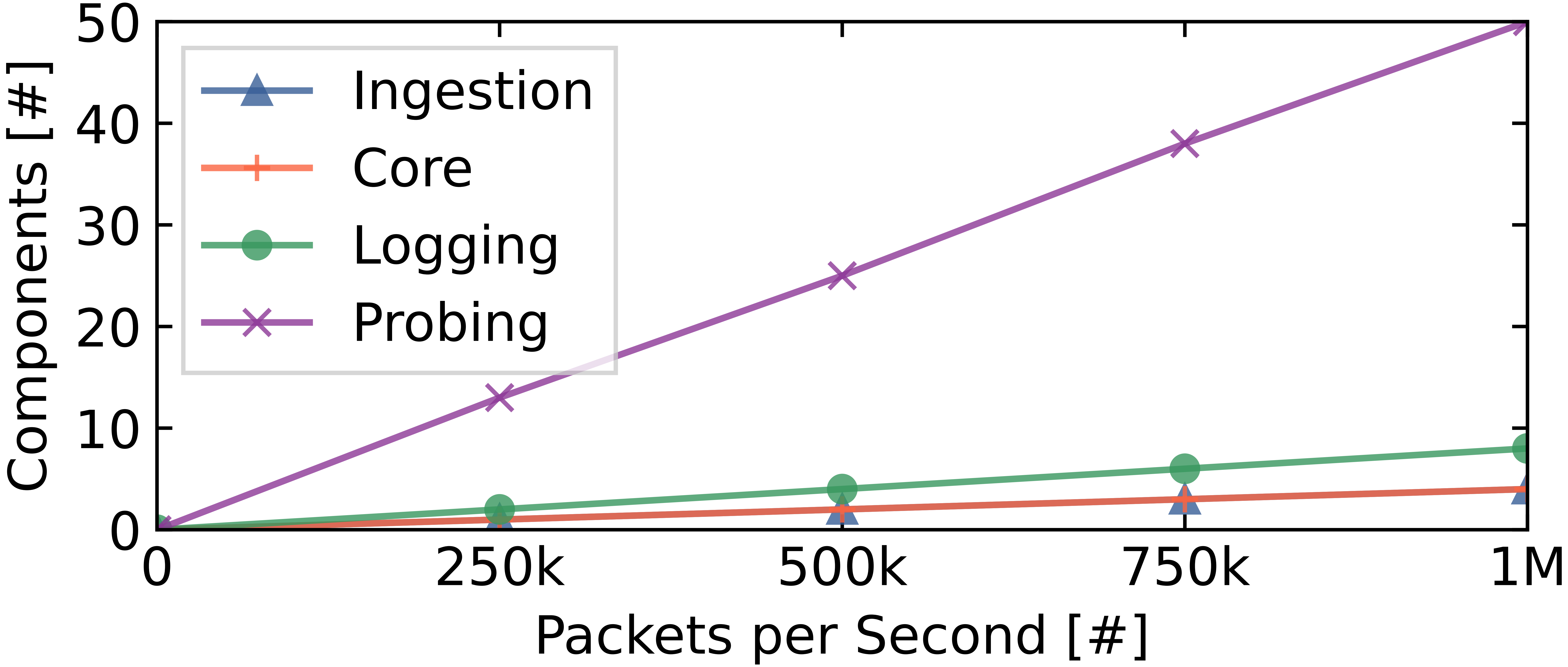
Architecture of Spoki



Architecture of Spoki



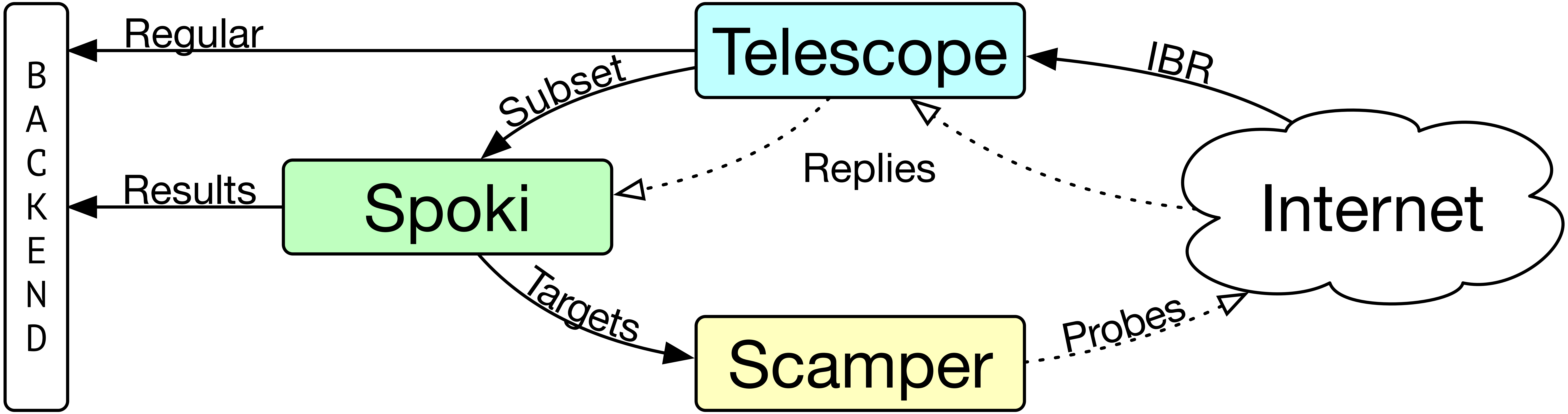
Scalability Measurements



Scales to /8: tested with up to 1M pps

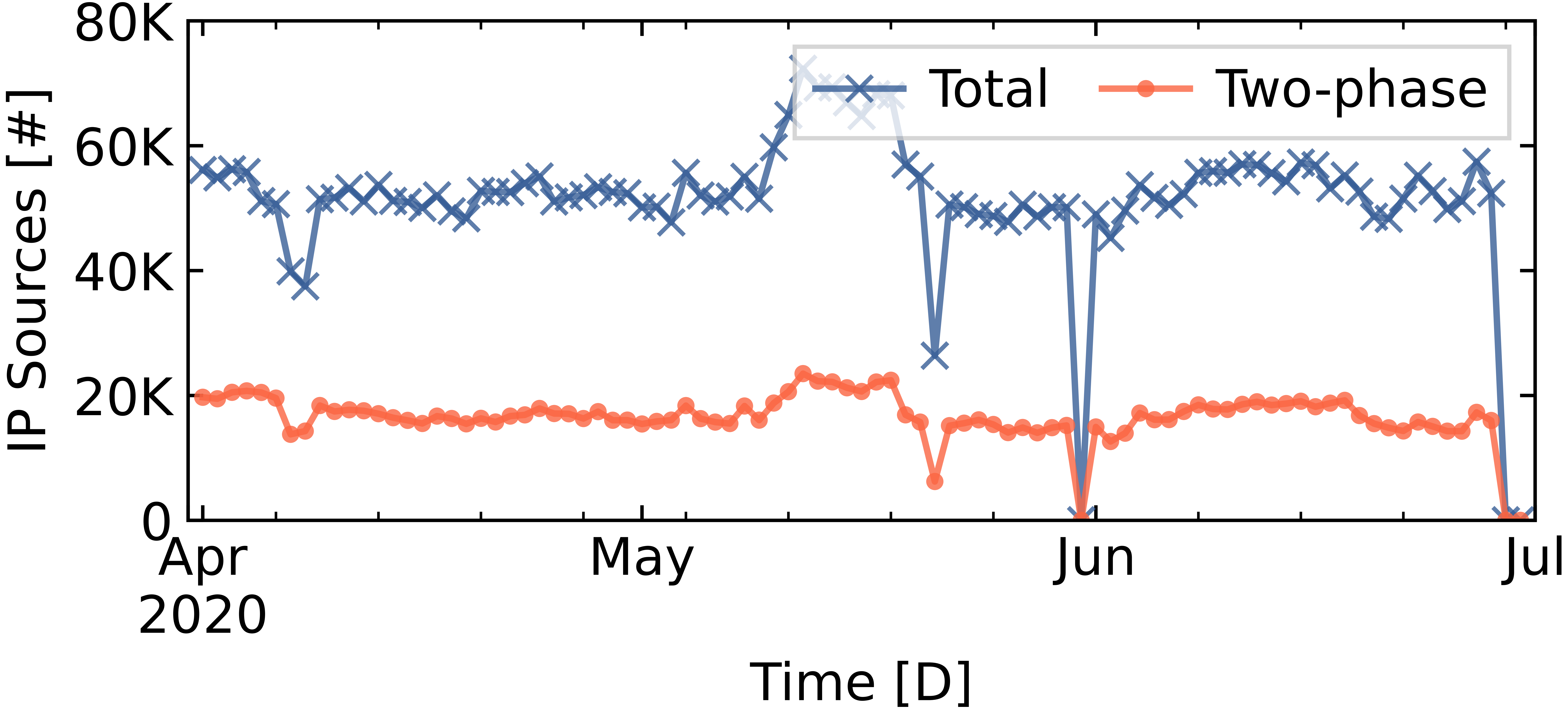
Spoki Deployment in a Reactive Telescope

- Data from two /24 networks in the US & EU
- Previously dark IP space that is not part of an active network
- Exclude well-known scanners from the analysis: 1.2% two-phase, 8.4% one-phase



Share of Two-phase Sources

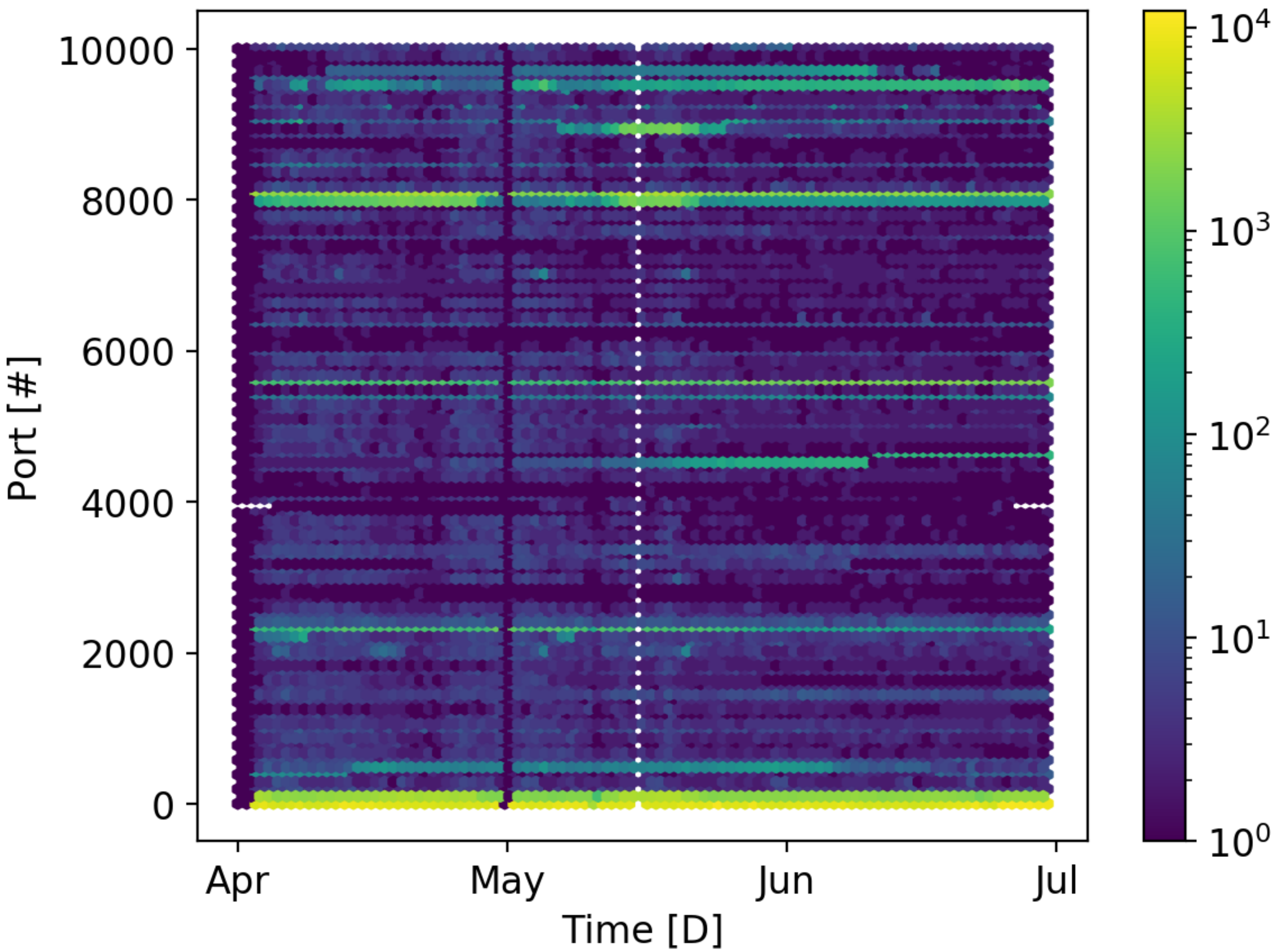
About 30% of sources send two-phase events each day.



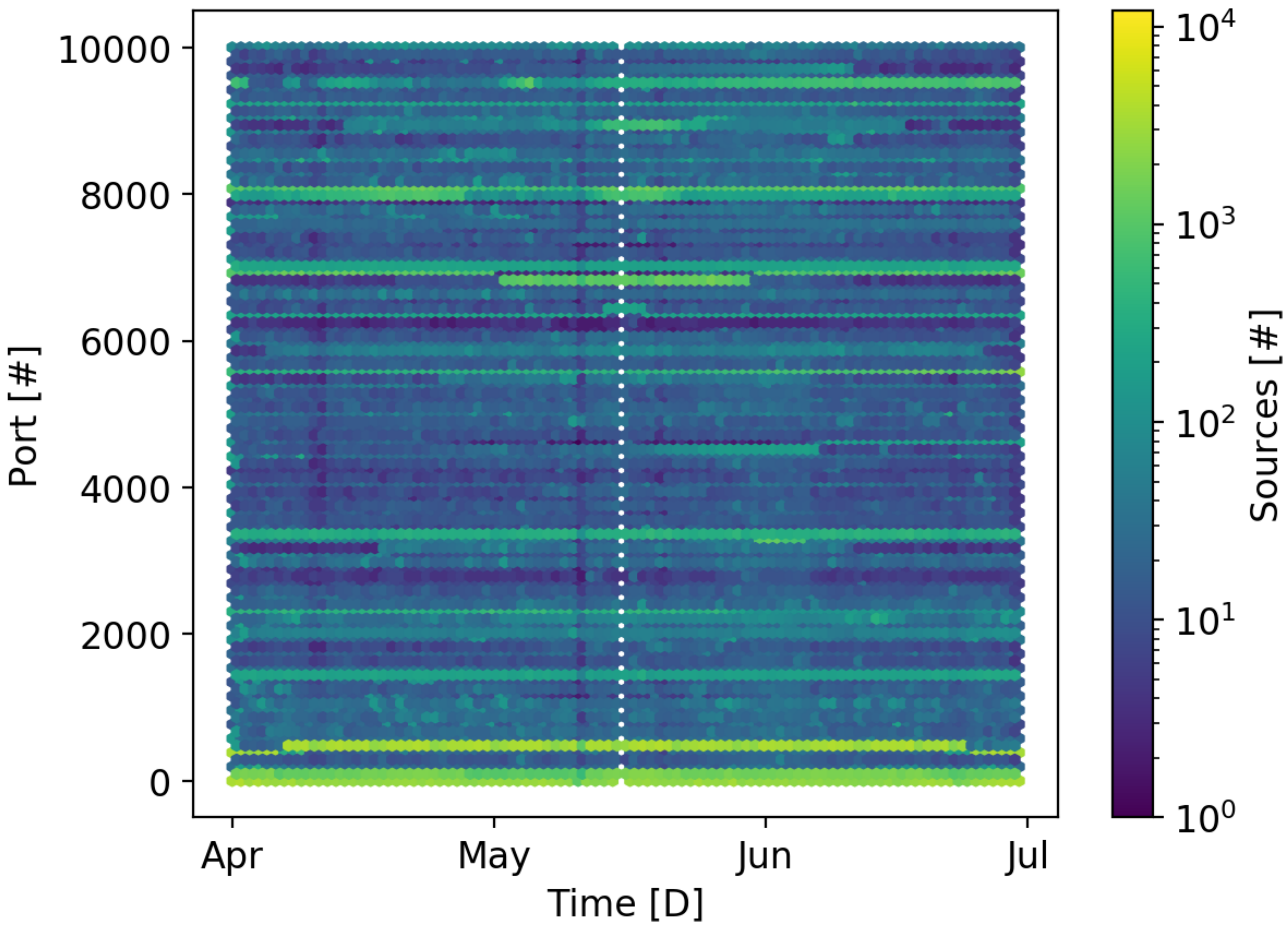
Scanning Activities

Two-phase scanners are more targeted than one-phase scanners.

Two-phase



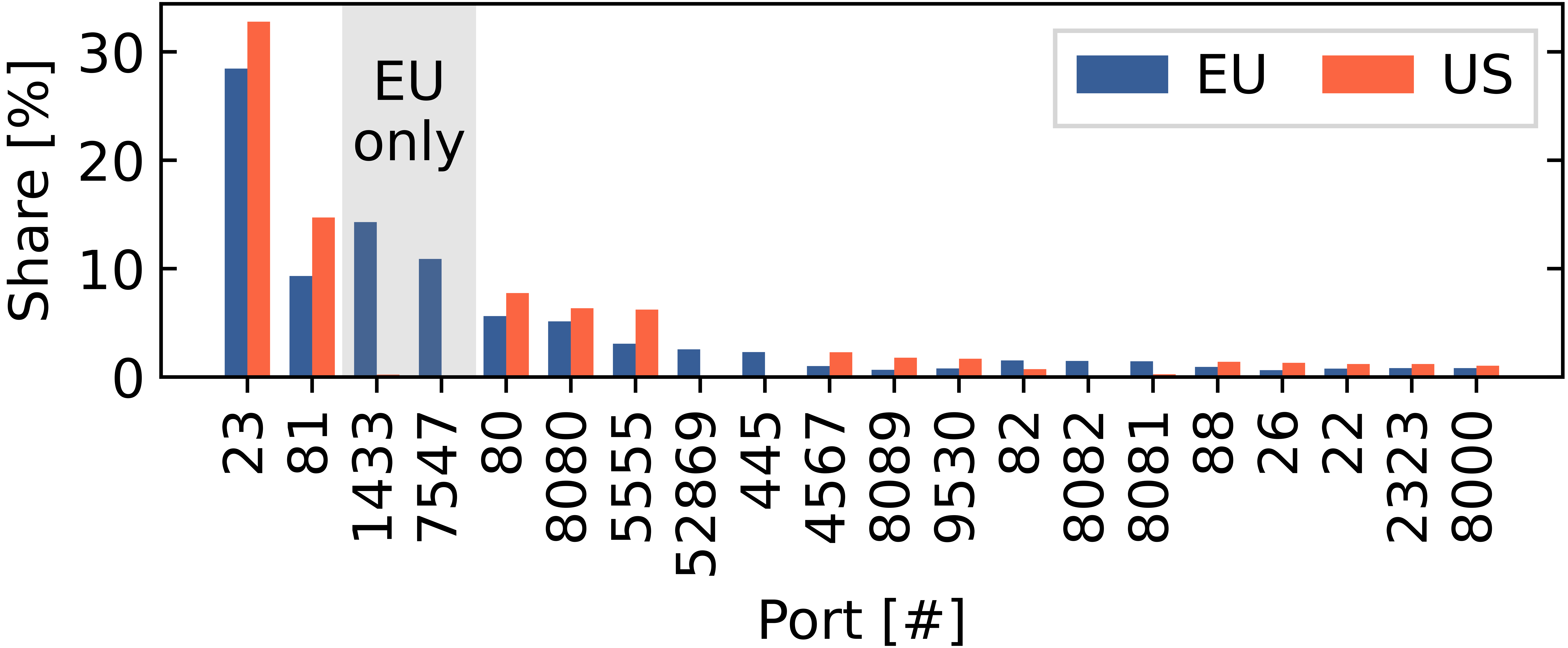
One-phase



Data is from the UCSD network telescope.

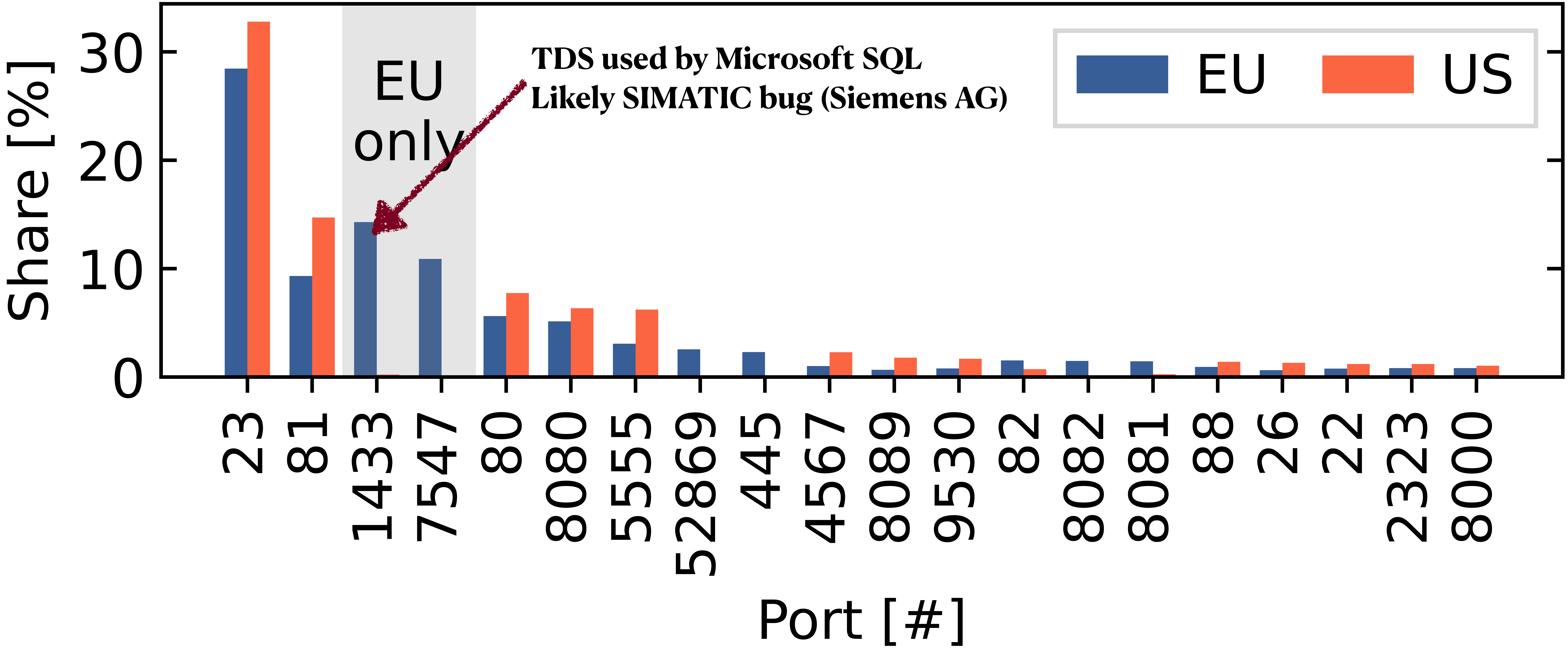
Targeted Ports

Two ports are scanned exclusively in the EU.



Targeted Ports

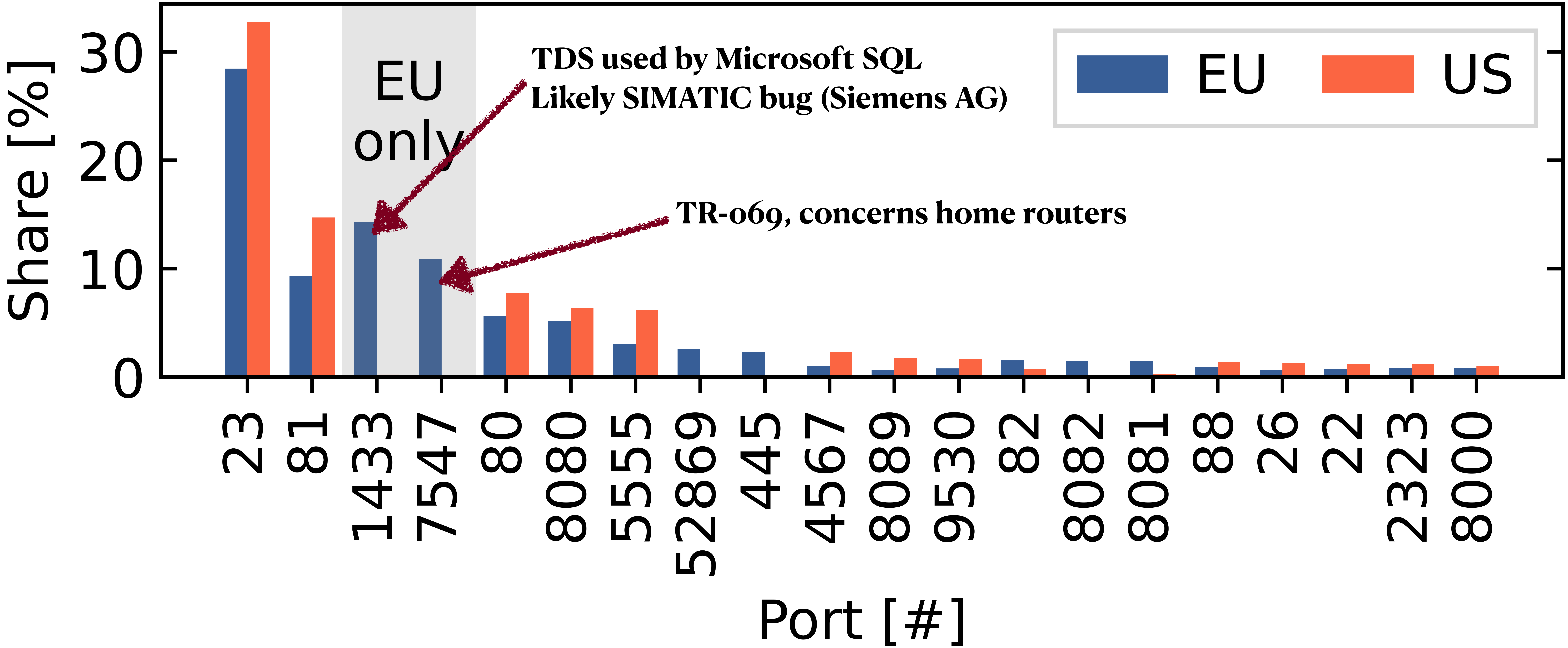
Two ports are scanned exclusively in the EU.



TDS used by Microsoft SQL
Likely SIMATIC bug (Siemens AG)

Targeted Ports

Two ports are scanned exclusively in the EU.



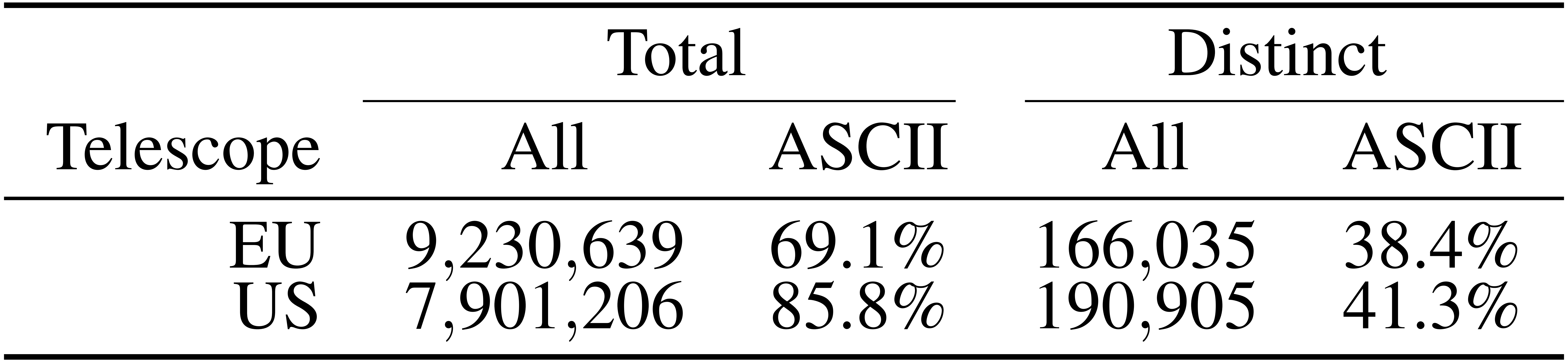
TDS used by Microsoft SQL

Likely SIMATIC bug (Siemens AG)

TR-069, concerns home routers

TCP Payloads

- Spoki accepts connections and collects ACK packets for a few seconds
- These payloads are not available in a traditional telescopes
- More than half of the payloads (in volume) are ASCII-decodable



	Total		Distinct	
Telescope	All	ASCII	All	ASCII
EU	9,230,639	69.1%	166,035	38.4%
US	7,901,206	85.8%	190,905	41.3%

The Maliciousness of Payloads

Semi-Manual

- Reveals several malicious payloads:

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

- Systematic approach needed to asses IPs:
  Query Threat Intelligence Provider

The Maliciousness of Payloads

Semi-Manual

- Reveals several malicious payloads:

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

- Systematic approach needed to asses IPs: Query Threat Intelligence Provider

GreyNoise

- Classifies IPs into malicious, benign, and unknown
- Share of malicious events:

	Two-phase	All
EU	56 %	38 %
US	70 %	35 %

- Two-phase events have a high share of malicious sources

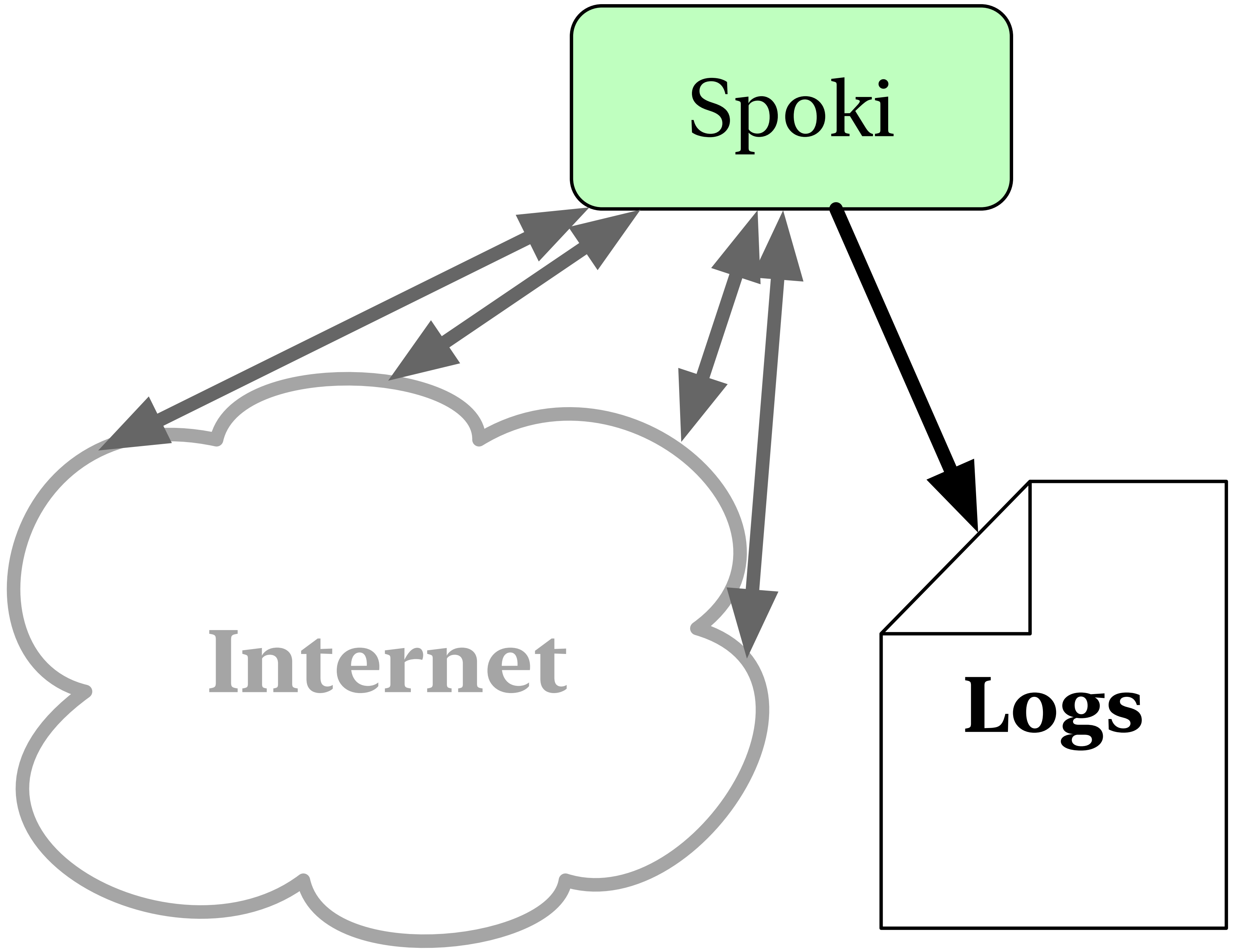
Shell Scripts & Malware Acquisition

- Some HTTP payloads include shell scripts, e.g.:

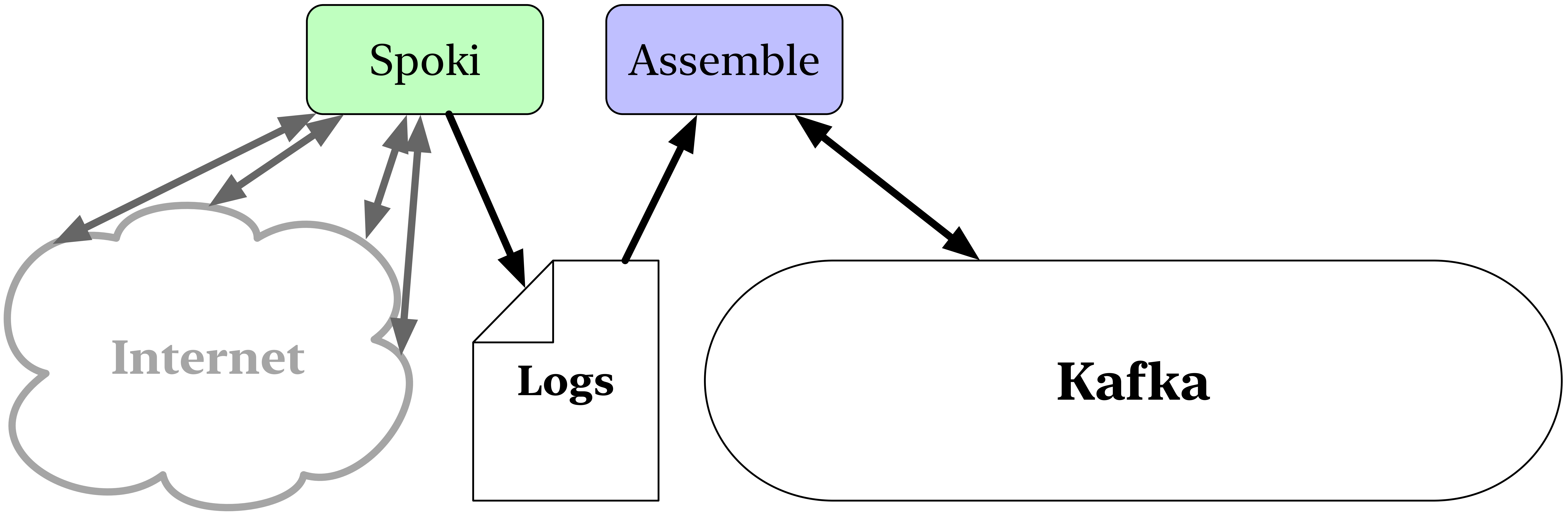
```
cd /tmp; rm -rf *;
  wget http://IPv4/arm7;
    chmod 777 arm7; ./arm7 rep.arm7
```

- Spoki can identify these snippets and download the malware

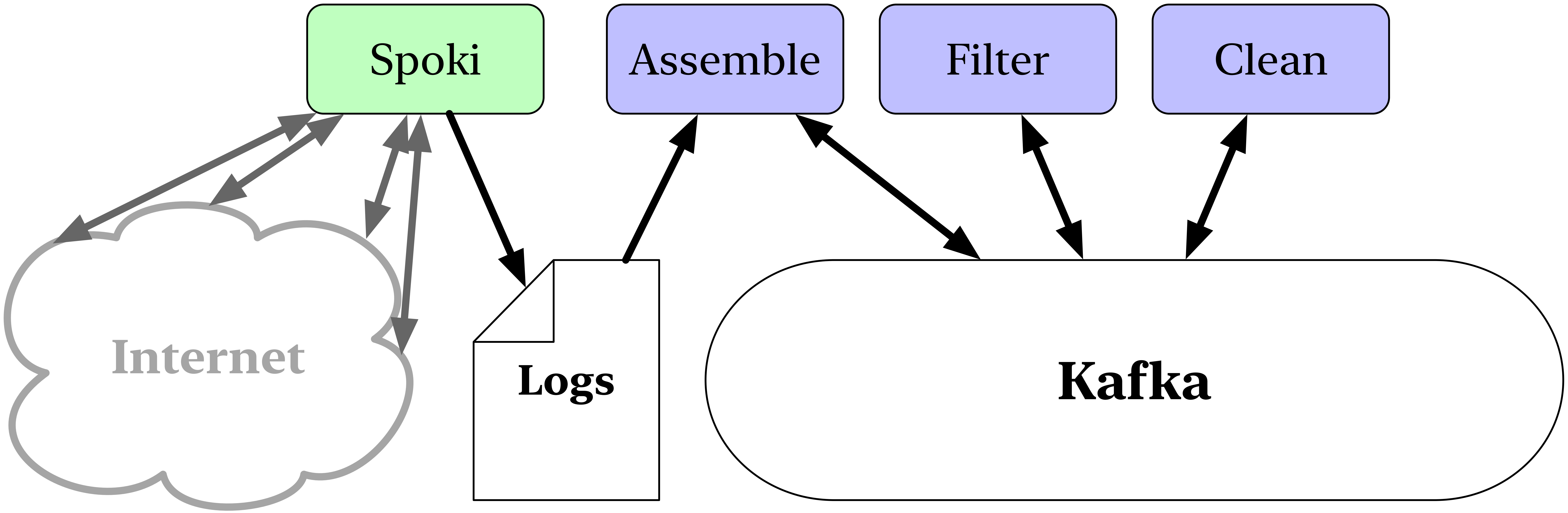
Malware Collection in Practice



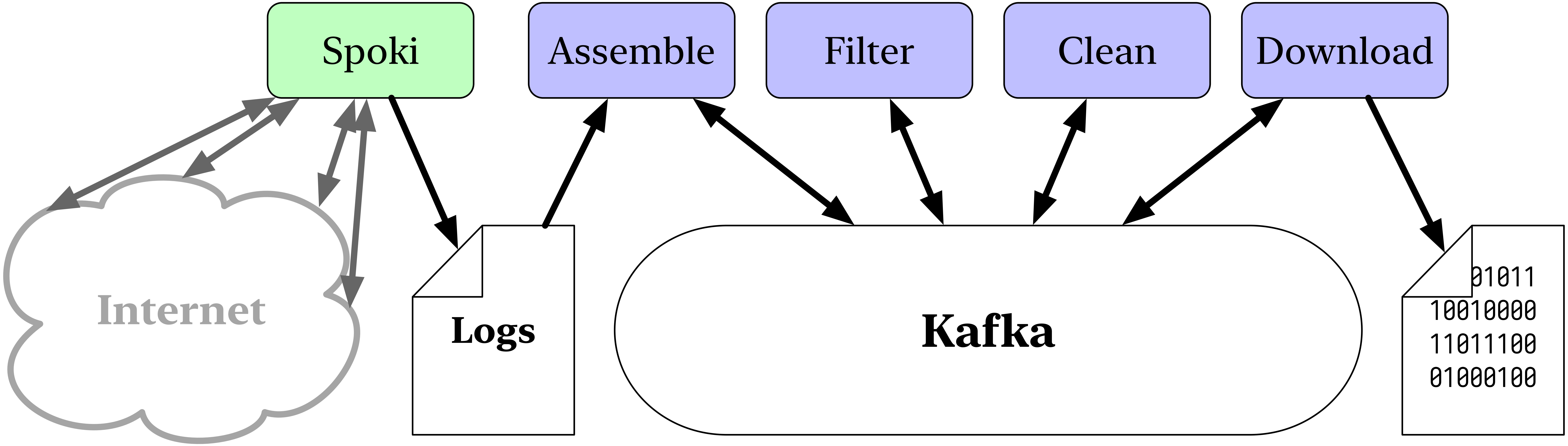
Malware Collection in Practice



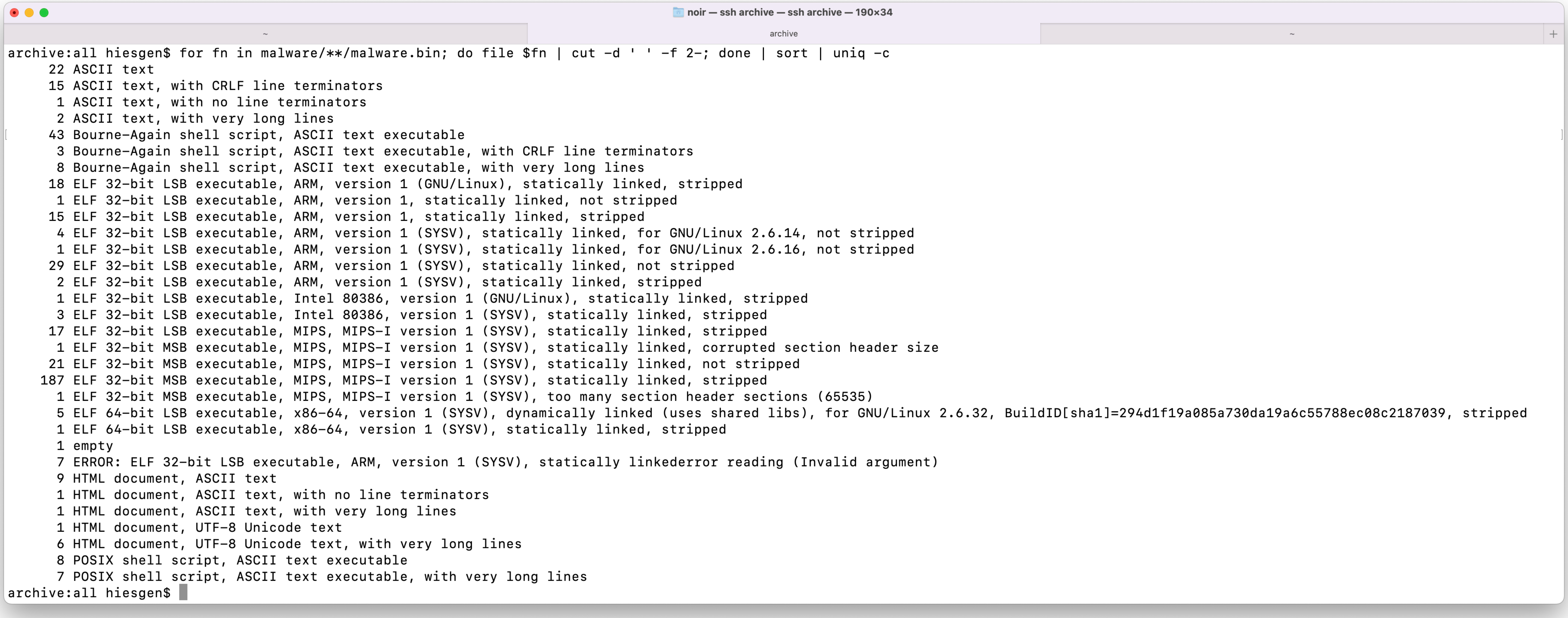
Malware Collection in Practice



Malware Collection in Practice



What did we find?

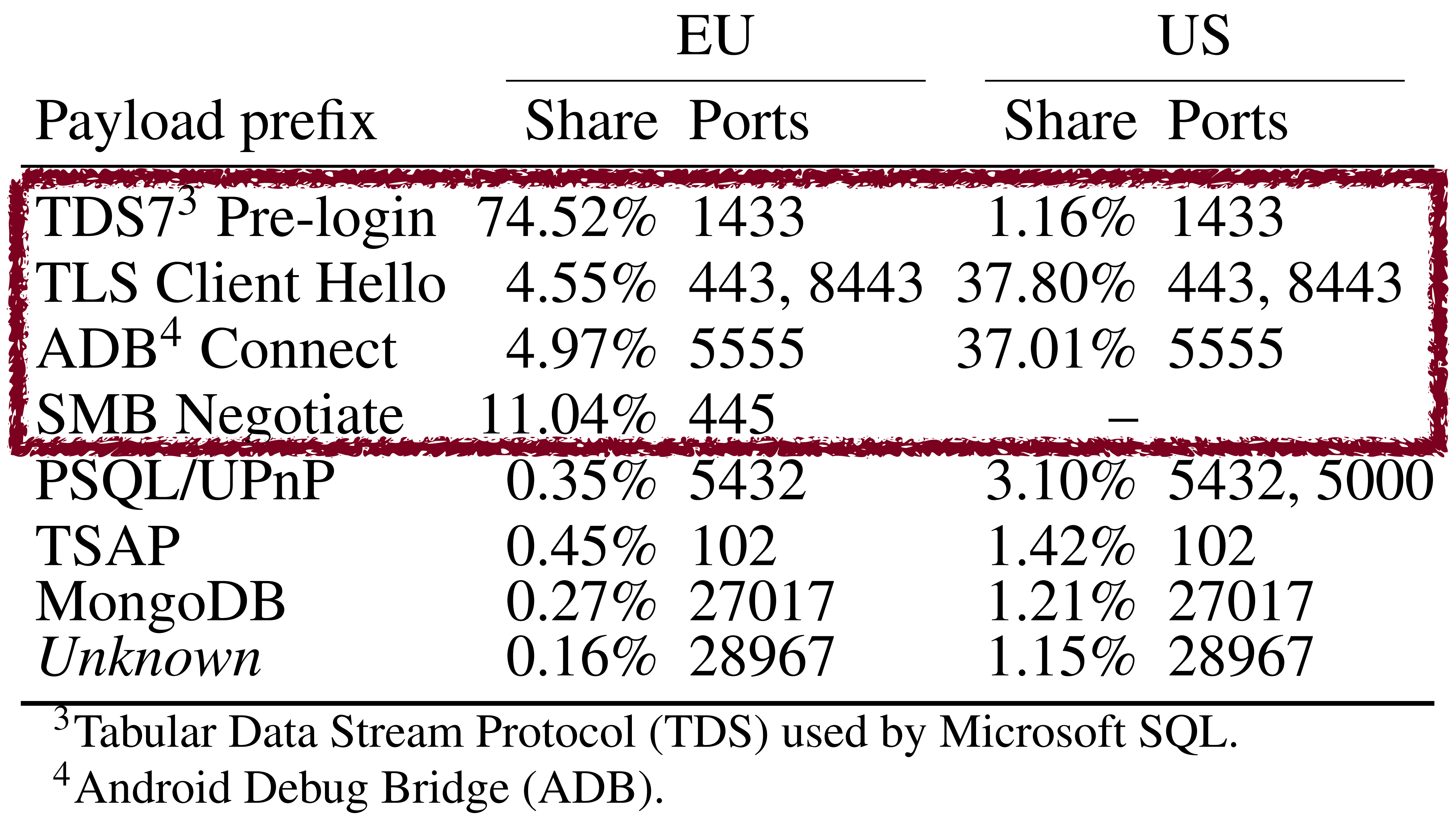


```
archive:all hiesgen$ for fn in malware/**/malware.bin; do file $fn | cut -d ' ' -f 2-; done | sort | uniq -c
     22 ASCII text
     15 ASCII text, with CRLF line terminators
      1 ASCII text, with no line terminators
      2 ASCII text, with very long lines
     43 Bourne-Again shell script, ASCII text executable
      3 Bourne-Again shell script, ASCII text executable, with CRLF line terminators
      8 Bourne-Again shell script, ASCII text executable, with very long lines
     18 ELF 32-bit LSB executable, ARM, version 1 (GNU/Linux), statically linked, stripped
      1 ELF 32-bit LSB executable, ARM, version 1, statically linked, not stripped
     15 ELF 32-bit LSB executable, ARM, version 1, statically linked, stripped
      4 ELF 32-bit LSB executable, ARM, version 1 (SYSV), statically linked, for GNU/Linux 2.6.14, not stripped
      1 ELF 32-bit LSB executable, ARM, version 1 (SYSV), statically linked, for GNU/Linux 2.6.16, not stripped
     29 ELF 32-bit LSB executable, ARM, version 1 (SYSV), statically linked, not stripped
      2 ELF 32-bit LSB executable, ARM, version 1 (SYSV), statically linked, stripped
      1 ELF 32-bit LSB executable, Intel 80386, version 1 (GNU/Linux), statically linked, stripped
      3 ELF 32-bit LSB executable, Intel 80386, version 1 (SYSV), statically linked, stripped
     17 ELF 32-bit LSB executable, MIPS, MIPS-I version 1 (SYSV), statically linked, stripped
      1 ELF 32-bit MSB executable, MIPS, MIPS-I version 1 (SYSV), statically linked, corrupted section header size
     21 ELF 32-bit MSB executable, MIPS, MIPS-I version 1 (SYSV), statically linked, not stripped
    187 ELF 32-bit MSB executable, MIPS, MIPS-I version 1 (SYSV), statically linked, stripped
      1 ELF 32-bit MSB executable, MIPS, MIPS-I version 1 (SYSV), too many section header sections (65535)
      5 ELF 64-bit LSB executable, x86-64, version 1 (SYSV), dynamically linked (uses shared libs), for GNU/Linux 2.6.32, BuildID[sha1]=294d1f19a085a730da19a6c55788ec08c2187039, stripped
      1 ELF 64-bit LSB executable, x86-64, version 1 (SYSV), statically linked, stripped
      1 empty
      7 ERROR: ELF 32-bit LSB executable, ARM, version 1 (SYSV), statically linkederror reading (Invalid argument)
      9 HTML document, ASCII text
      1 HTML document, ASCII text, with no line terminators
      1 HTML document, ASCII text, with very long lines
      1 HTML document, UTF-8 Unicode text
      6 HTML document, UTF-8 Unicode text, with very long lines
      8 POSIX shell script, ASCII text executable
      7 POSIX shell script, ASCII text executable, with very long lines
archive:all hiesgen$ 
```

- **Spoki detected 15% of the hashes earlier than VirusTotal** (26% benign, 59% old)

Geographical Scanning Locality

- Ports 1433 & 7547 are nearly exclusively visible in the EU
- Payloads to 5555 and 443 take a much higher share in the US



	EU		US	
Payload prefix	Share	Ports	Share	Ports
TDS7[3] Pre-login	74.52%	1433	1.16%	1433
TLS Client Hello	4.55%	443, 8443	37.80%	443, 8443
ADB[4] 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

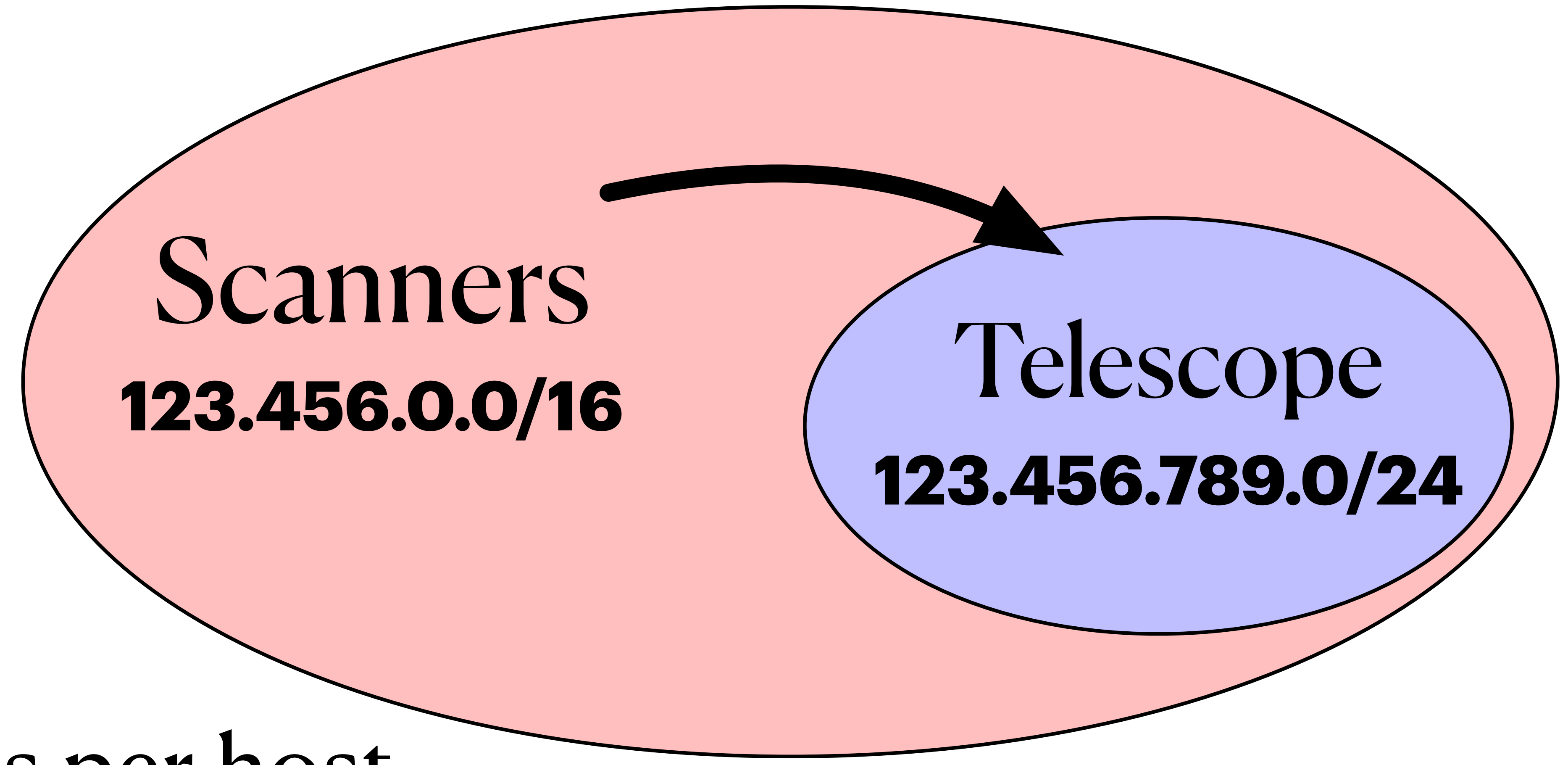
[3] Tabular Data Stream Protocol (TDS) used by Microsoft SQL.

[4] Android Debug Bridge (ADB).

Collected HEX Payloads

Topological Scanning Locality

- Six of the top-ten source prefixes in the EU share a /16 with our /24 vantage point
  - Geographic origins in UA, PL, and RU
  - A similar locality cannot be observed in the US

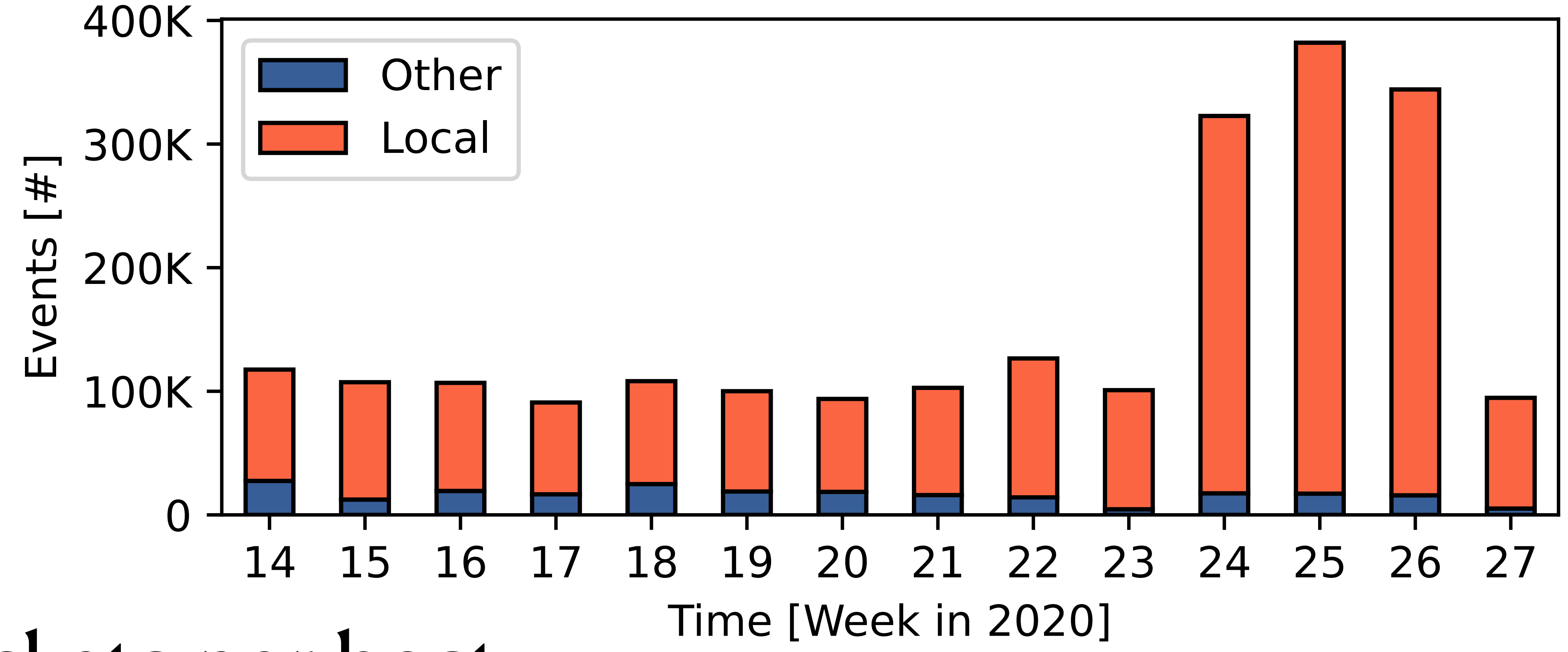


- Crosscheck (sampled) traffic at a European IXP
  - Local, irregular SYNs in 370 prefixes with about 150 packets per host
  - Local traffic targets 23, 7547, 8291 while non-local traffic targets 80, 443, 23

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

Topological Scanning Locality

- Six of the top-ten source prefixes in the EU share a /16 with our /24 vantage point
  - Geographic origins in UA, PL, and RU
  - A similar locality cannot be observed in the US



- Crosscheck (sampled) traffic at a European IXP
  - Local, irregular SYNs in 370 prefixes with about 150 packets per host
  - Local traffic targets 23, 7547, 8291 while non-local traffic targets 80, 443, 23

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

Takeaways

- Spoki: Designed a highly scalable reactive telescope
- Irregular SYNs dominate SYNs on the Internet: ~75%
- Two-phase scans
  - … are highly focused
  - … are used for malicious activities (GN: 50-70% malicious sources)
- Two-phase events follow locality patterns, both geographically and topologically