Illuminating Large-Scale IPv6 Scanning in the Internet

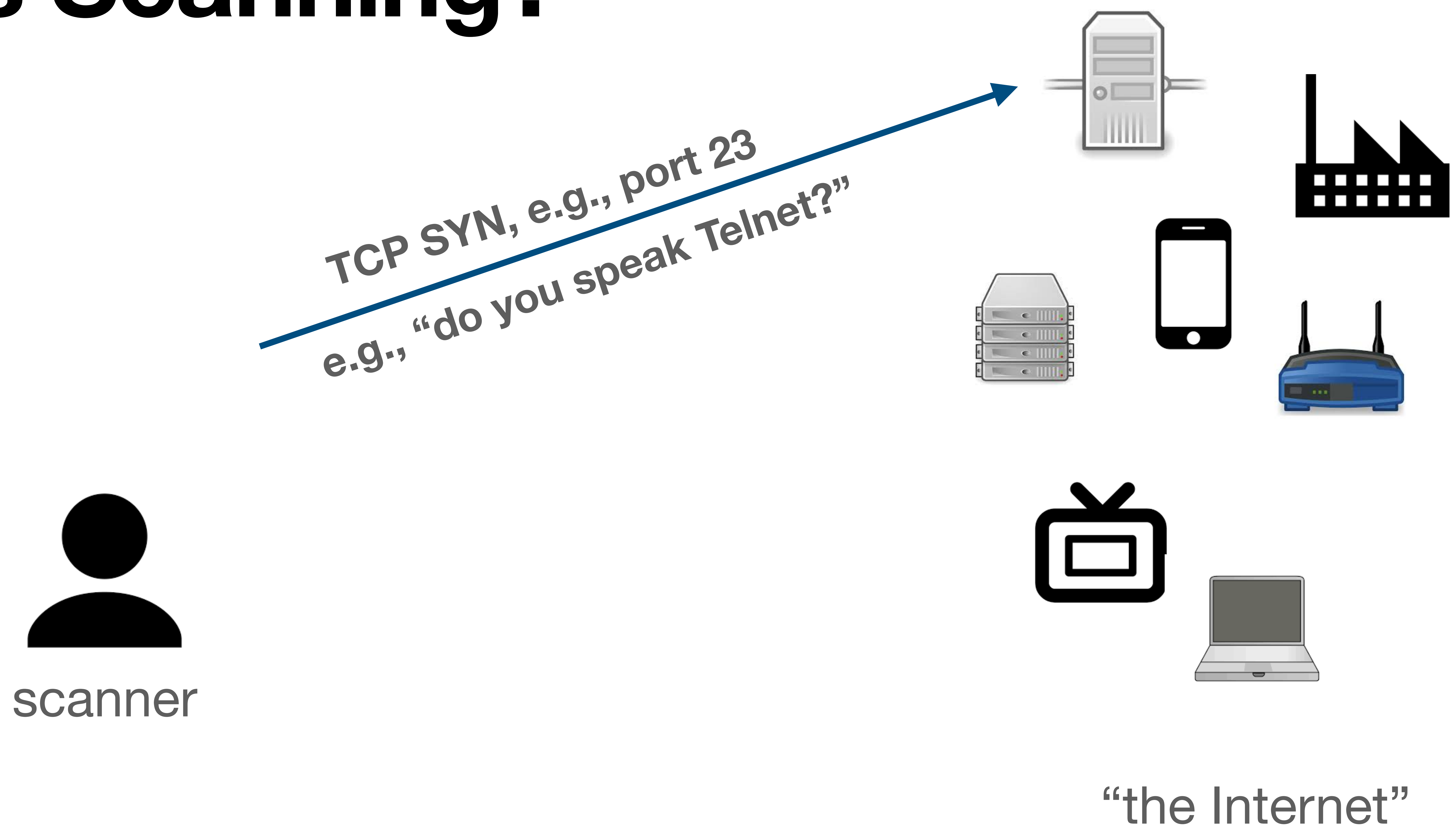
Philipp Richter, Oliver Gasser, and Arthur Berger

ACM Internet Measurement Conference 2022
Nice, France

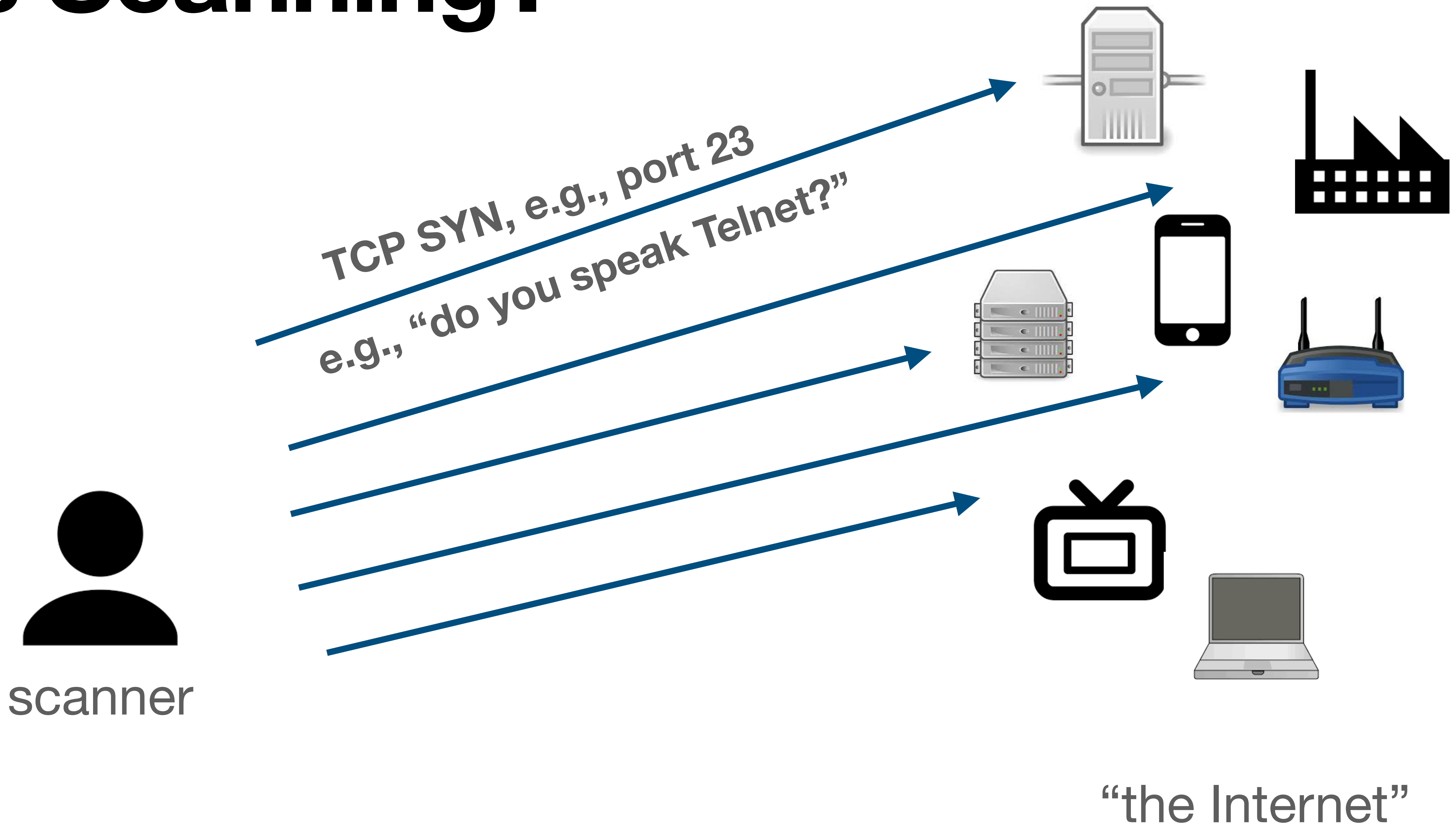
Akamai

max planck institut
informatik

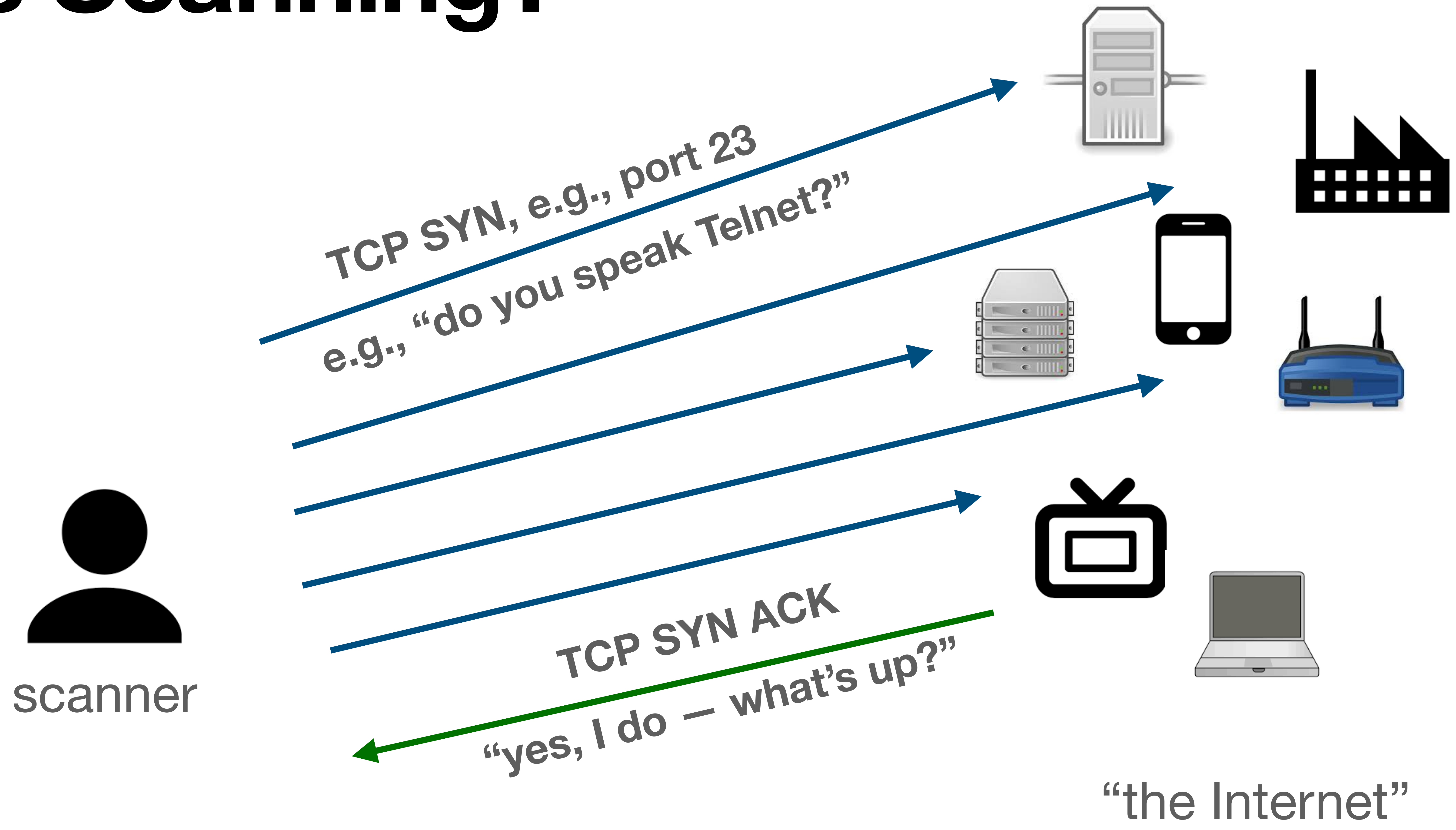
What is Scanning?



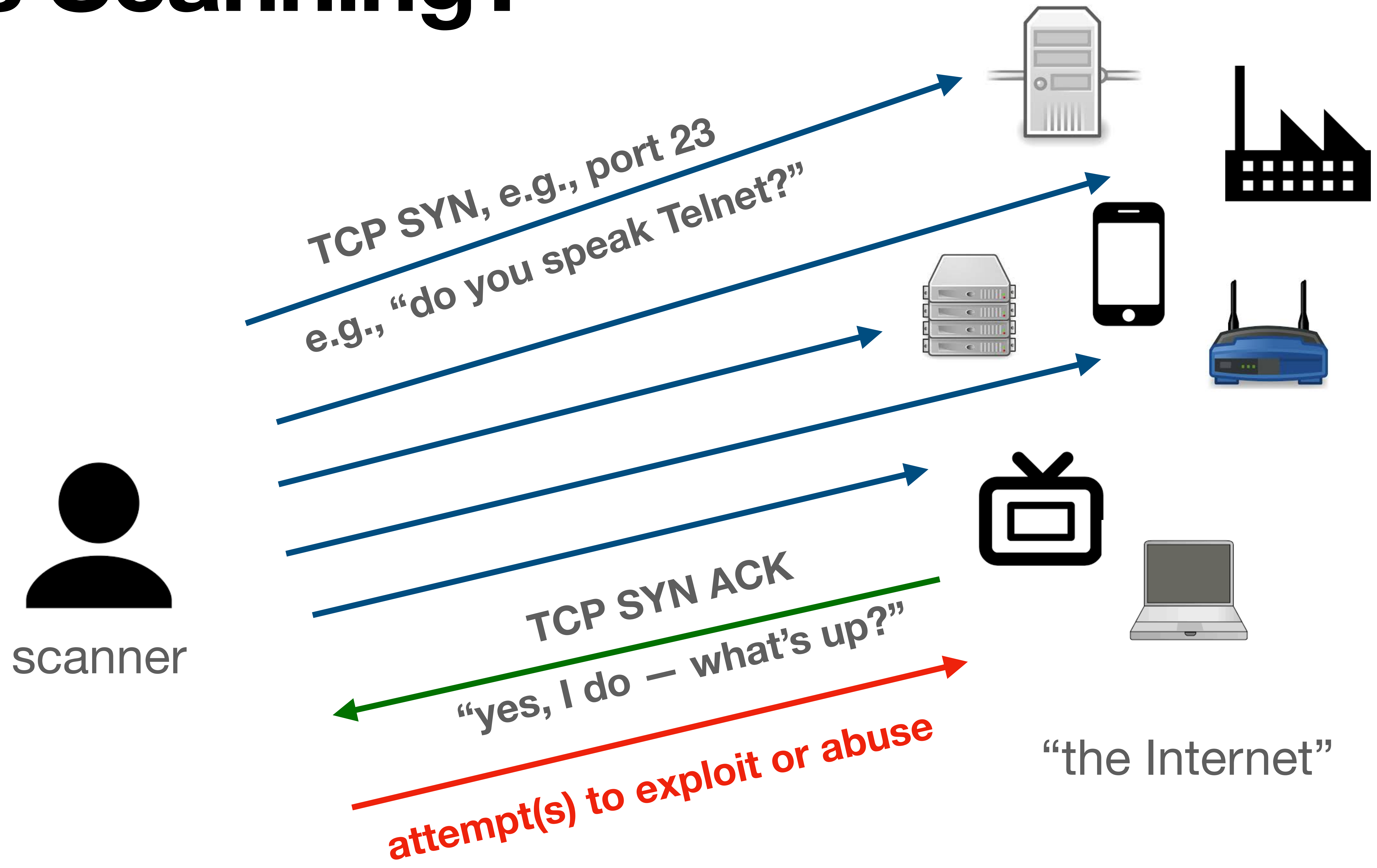
What is Scanning?



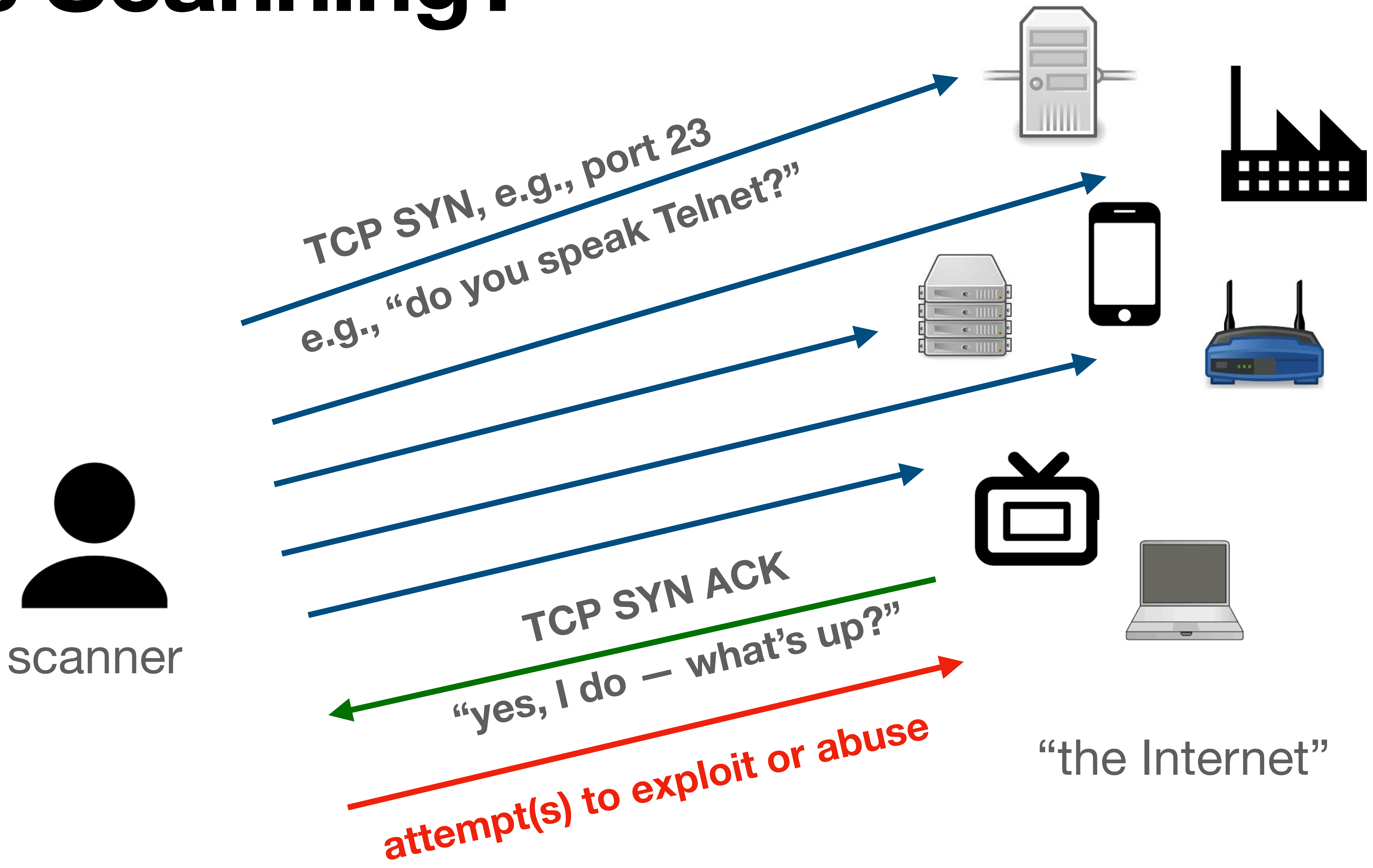
What is Scanning?



What is Scanning?



What is Scanning?



Scanning is key for cyberattacks.

Scanning in IPv4

- About 4 billion target addresses
  e.g., `198.51.100.17`
- Full scan in <1 hour
- Scan detection readily possible
  (e.g., using darknets)**
- Millions of monthly active scanners

** with limitations

Scanning in IPv4

- About 4 billion target addresses
  e.g., `198.51.100.17`
- Full scan in <1 hour
- Scan detection readily possible
  (e.g., using darknets)**
- Millions of monthly active scanners

Scanning in IPv6

- About 10^{38} target addresses
  e.g., `2001:db8:86e7:637:106c:d7dc:248:4a5d`
- Trillions of years needed for full scan
- Detection not readily possible
  (need vantage points!)
- Extent of active scanning unknown

** with limitations

Scanning in IPv4

- About 4 billion target addresses
  e.g., `198.51.100.17`
- Full scan in <1 hour
- Scan detection readily possible
  (e.g., using darknets)**
- Millions of monthly active scanners

Scanning in IPv6

- About 10^{38} target addresses
  e.g., `2001:db8:86e7:637:106c:d7dc:248:4a5d`
- Trillions of years needed for full scan
- Detection not readily possible
  (need vantage points!)
- Extent of active scanning unknown

What's going on in the IPv6 space?

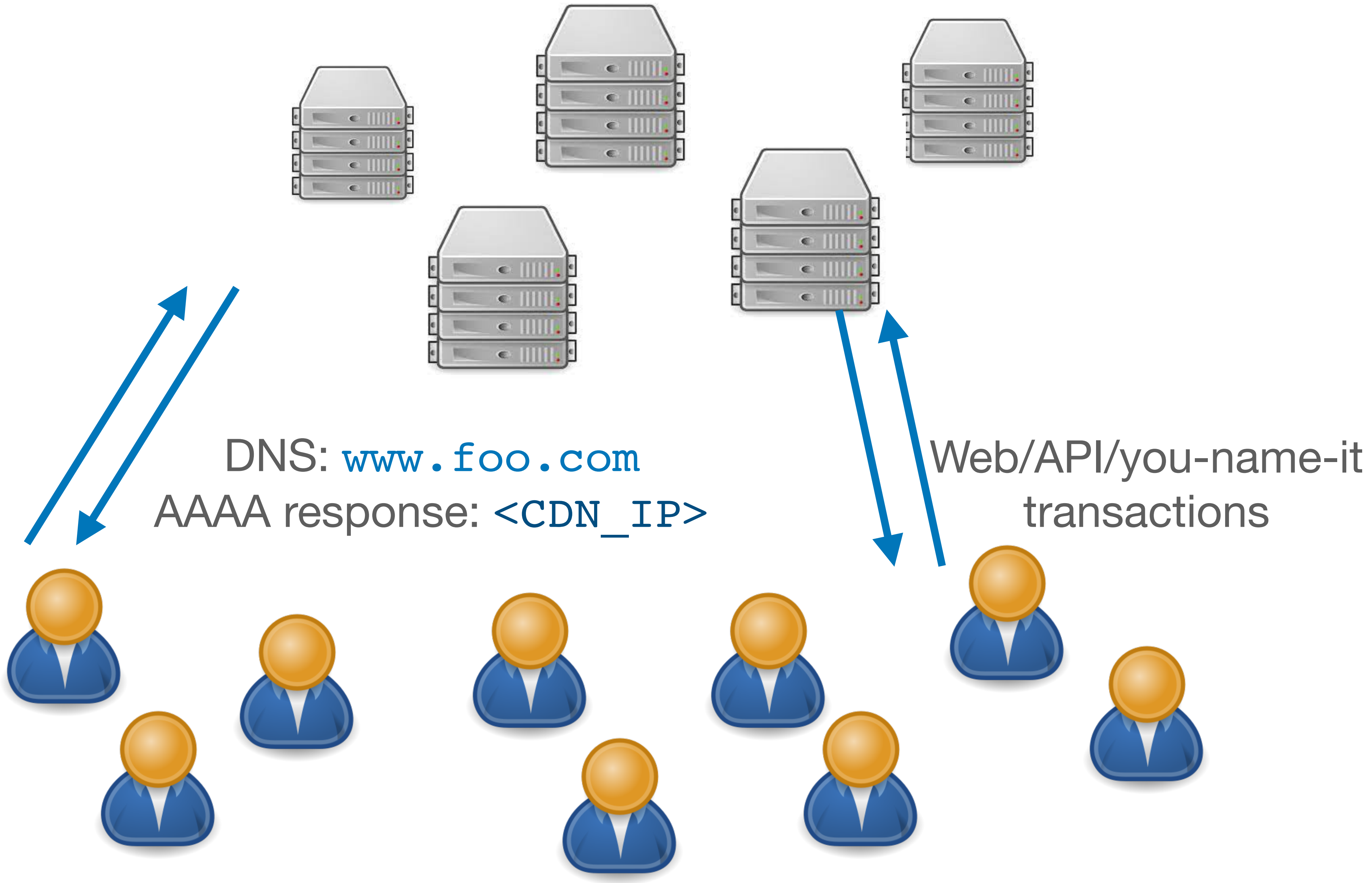
** with limitations

First Longitudinal Study of Large-Scale IPv6 Scans

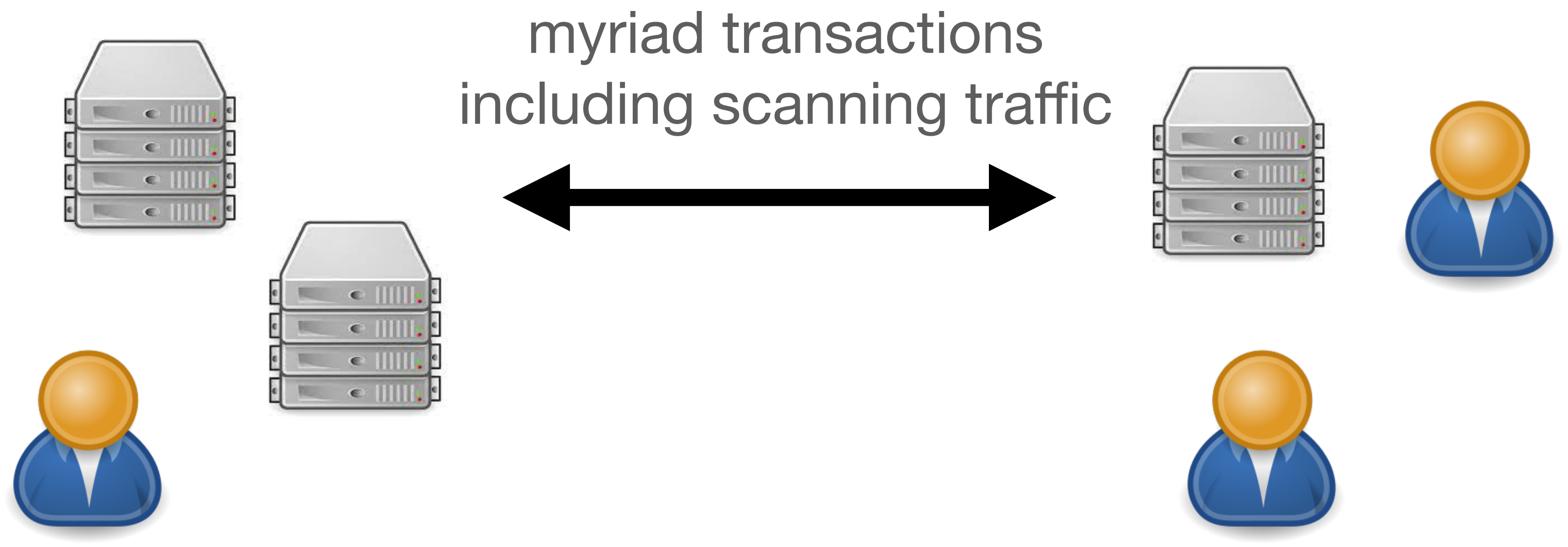
- 15 months of firewall logs of some 200,000+ CDN servers
- Double-check with publicly available traffic traces (MAWI)

First Longitudinal Study of Large-Scale IPv6 Scans

- 15 months of firewall logs of some 200,000+ CDN servers
- Double-check with publicly available traffic traces (MAWI)



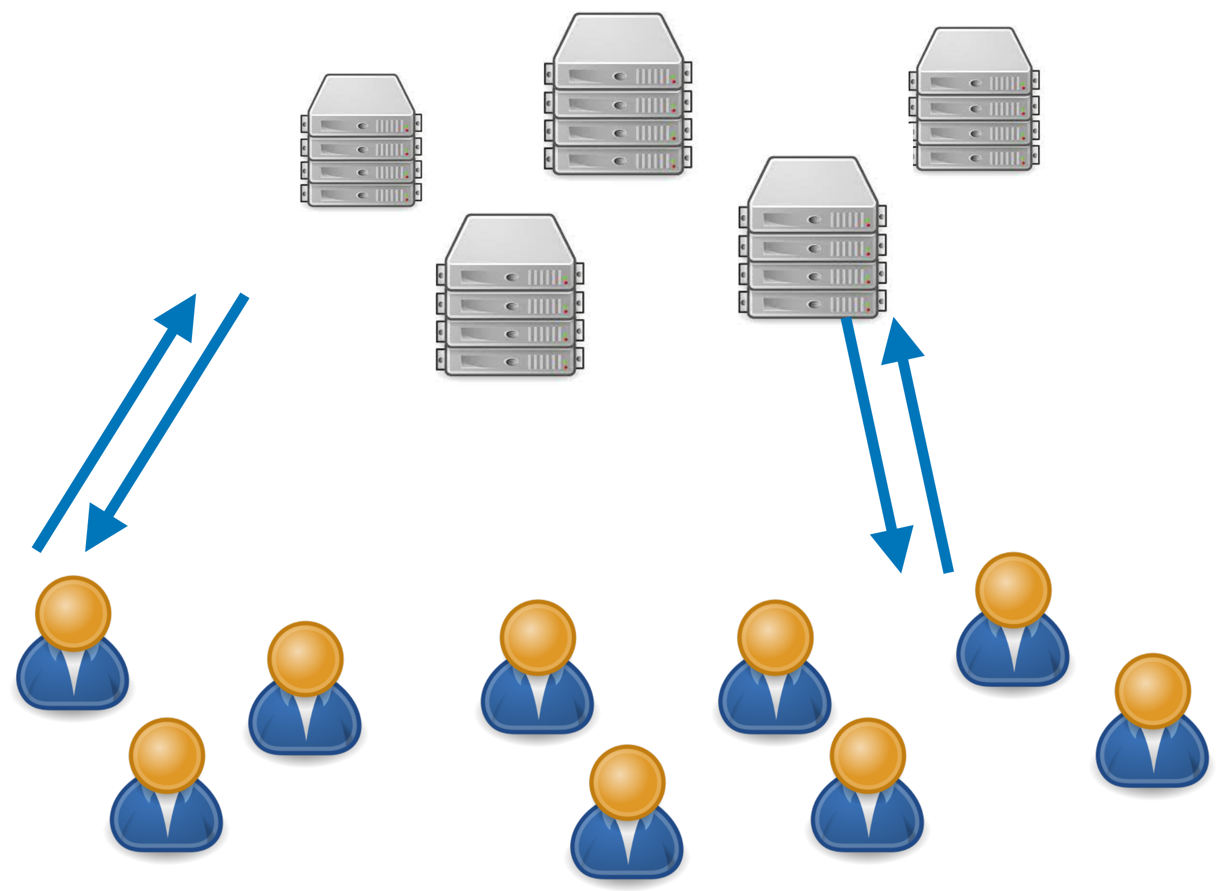
CDN firewall logs:
Target address exposure via DNS, among others.



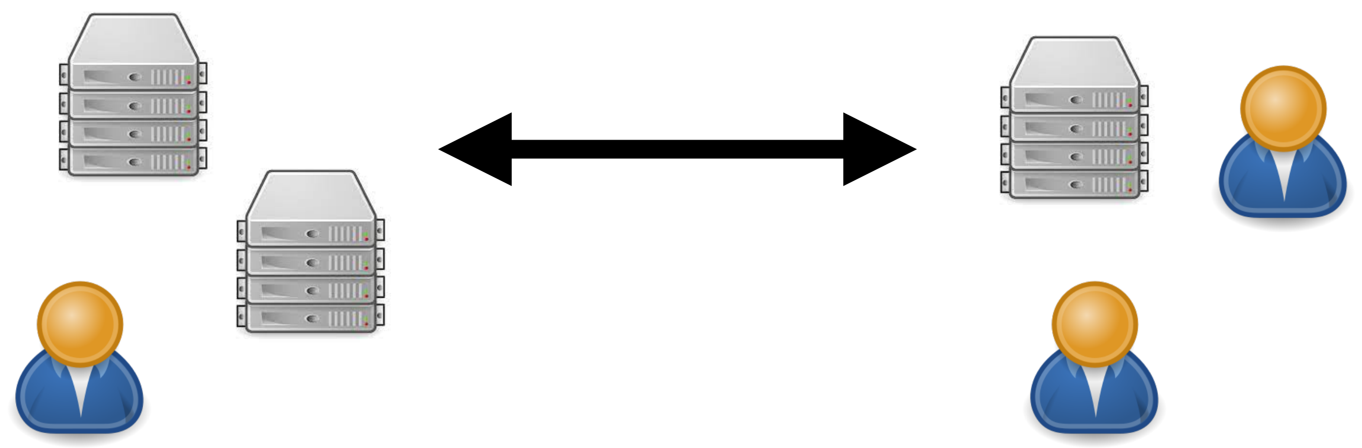
MAWI passive traces:
capture on-the-wire traffic, including scanning

First Longitudinal Study of Large-Scale IPv6 Scans

- 15 months of firewall logs of some 200,000+ CDN servers
- Double-check with publicly available traffic traces (MAWI)



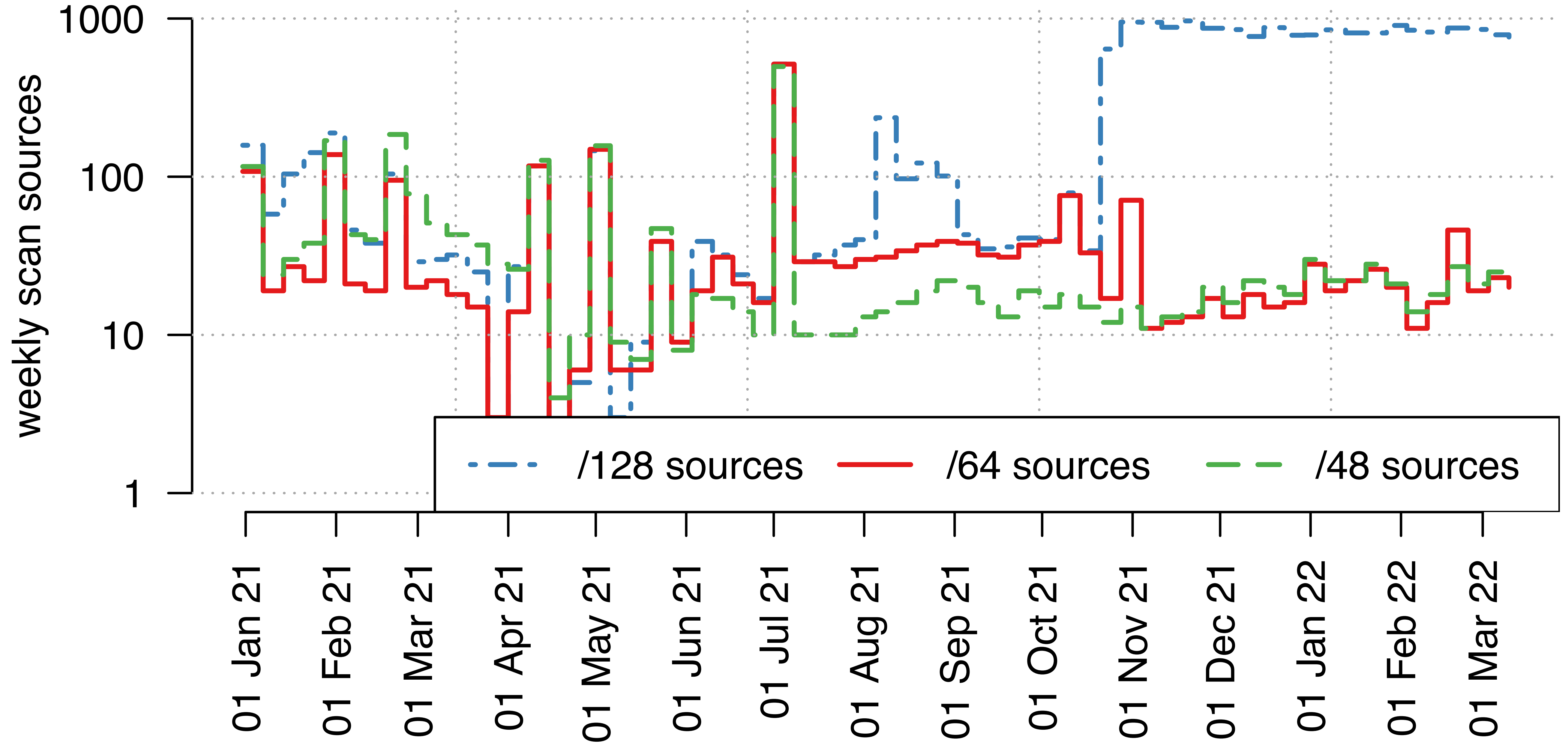
CDN firewall logs:
Target address exposure via DNS, among others.



MAWI passive traces:
capture on-the-wire traffic, including scanning

Large-Scale IPv6 Scans:
Sources that target at least 100 DST IPs in either vantage point.

IPv6 Scan Sources over Time



IPv6 is now actively scanned.
We find between ~10 and ~100 active weekly sources.

Top IPv6 Scan Source Networks

rank	AS type	packets	scan sources /48s	/64s	/128s
#1	Datacenter (CN)	839M (39.2%)	1	1	1
#2	Datacenter (CN)	744M (34.8%)	1	1	5
#3	Cybersecurity (US)	275M (12.9%)	1	1	12
#4	Cloud (US/global)	78M (3.7%)	2	2	512
#5	Cloud (DE)	48M (2.3%)	3	59	59
#6	Cloud (US/global)	45M (2.1%)	10	15	205
#7	Cloud (US/global)	39M (1.8%)	9	9	123
#8	Cloud (CN)	30M (1.4%)	5	5	53
#9	Transit (global)	11M (0.5%)	1	2	956
#10	Cloud (CN)	10M (0.5%)	1	1	7
#11	Cloud (US/global)	4.7M (0.2%)	1	1	353
#12	Datacenter (CN)	3.1M (0.1%)	9	12	19
#13	ISP (VN)	2.5M (0.1%)	1	1	1
#14	Datacenter (CN)	1.6M ($\leq$ 0.1%)	1	1	2
#15	Research (DE)	1.1M ($\leq$ 0.1%)	1	1	1
#16	ISP (RU)	0.9M ($\leq$ 0.1%)	1	1	2
#17	University (DE)	0.8M ($\leq$ 0.1%)	1	1	2
#18	Cloud/Transit (DE)	0.6M ($\leq$ 0.1%)	1,092	1,057	1,057
#19	ISP (RU)	0.6M ($\leq$ 0.1%)	1	1	1
#20	University (DE)	0.5M ($\leq$ 0.1%)	1	1	1

Traffic heavily concentrated on datacenter/cloud ASes.

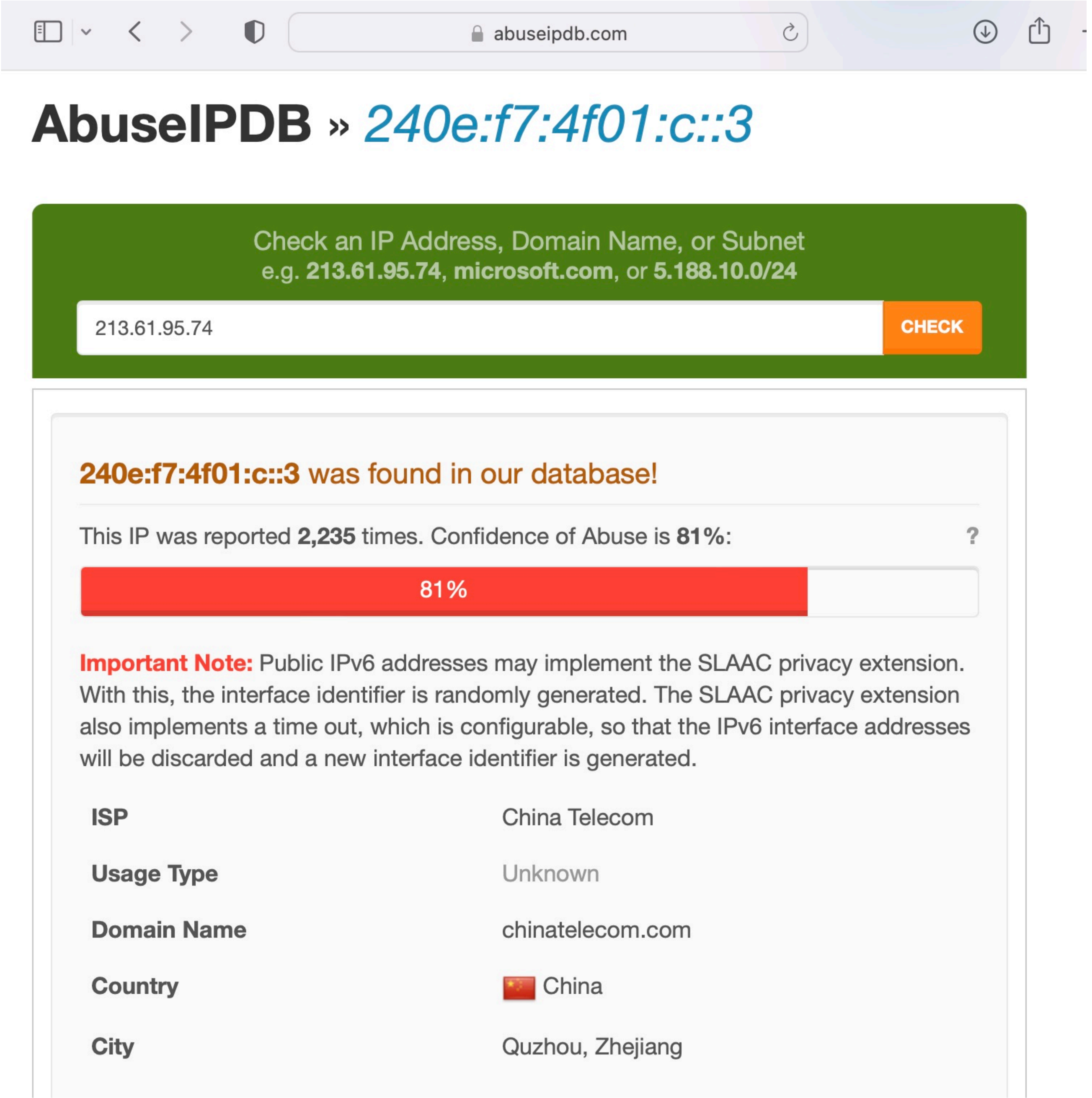
Top IPv6 Scan Source Networks

rank	AS type	packets	scan sources /48s	/64s	/128s
#1	Datacenter (CN)	839M (39.2%)	1	1	1
#2	Datacenter (CN)	744M (34.8%)	1	1	5
#3	Cybersecurity (US)	275M (12.9%)	1	1	12
#4	Cloud (US/global)	78M (3.7%)	2	2	512
#5	Cloud (DE)	48M (2.3%)	3	59	59
#6	Cloud (US/global)	45M (2.1%)	10	15	205
#7	Cloud (US/global)	39M (1.8%)	9	9	123
#8	Cloud (CN)	30M (1.4%)	5	5	53
#9	Transit (global)	11M (0.5%)	1	2	956
#10	Cloud (CN)	10M (0.5%)	1	1	7
#11	Cloud (US/global)	4.7M (0.2%)	1	1	353
#12	Datacenter (CN)	3.1M (0.1%)	9	12	19
#13	ISP (VN)	2.5M (0.1%)	1	1	1
#14	Datacenter (CN)	1.6M ($\leq$ 0.1%)	1	1	2
#15	Research (DE)	1.1M ($\leq$ 0.1%)	1	1	1
#16	ISP (RU)	0.9M ($\leq$ 0.1%)	1	1	2
#17	University (DE)	0.8M ($\leq$ 0.1%)	1	1	2
#18	Cloud/Transit (DE)	0.6M ($\leq$ 0.1%)	1,092	1,057	1,057
#19	ISP (RU)	0.6M ($\leq$ 0.1%)	1	1	1
#20	University (DE)	0.5M ($\leq$ 0.1%)	1	1	1

Traffic heavily concentrated on datacenter/cloud ASes.

Topmost Active IPv6 Scan Source

- Single most active source in **CDN firewall and passive MAWI trace!**
- Continually active for almost 2 years
- Scanning right now!
  (though changing ports targeted)
- Reported 1000s of times in open-source reputation data



Ports Targeted

- Majority of scans target *multiple* port numbers / services
- Behavior resembling that of general penetration testing as opposed to exploitation of specific vulnerabilities

Top IPv6 Scan Source Networks

rank	AS type	packets	scan sources /48s	/64s	/128s
#1	Datacenter (CN)	839M (39.2%)	1	1	1
#2	Datacenter (CN)	744M (34.8%)	1	1	5
#3	Cybersecurity (US)	275M (12.9%)	1	1	12
#4	Cloud (US/global)	78M (3.7%)	2	2	512
#5	Cloud (DE)	48M (2.3%)	3	59	59
#6	Cloud (US/global)	45M (2.1%)	10	15	205
#7	Cloud (US/global)	39M (1.8%)	9	9	123
#8	Cloud (CN)	30M (1.4%)	5	5	53
#9	Transit (global)	11M (0.5%)	1	2	956
#10	Cloud (CN)	10M (0.5%)	1	1	7
#11	Cloud (US/global)	4.7M (0.2%)	1	1	353
#12	Datacenter (CN)	3.1M (0.1%)	9	12	19
#13	ISP (VN)	2.5M (0.1%)	1	1	1
#14	Datacenter (CN)	1.6M (≤ 0.1%)	1	1	2
#15	Research (DE)	1.1M (≤ 0.1%)	1	1	1
#16	ISP (RU)	0.9M (≤ 0.1%)	1	1	2
#17	University (DE)	0.8M (≤ 0.1%)	1	1	2
#18	Cloud/Transit (DE)	0.6M (≤ 0.1%)	1,092	1,057	1,057
#19	ISP (RU)	0.6M (≤ 0.1%)	1	1	1
#20	University (DE)	0.5M (≤ 0.1%)	1	1	1

Major Challenge: Identifying and isolating scan sources.

Key Challenge: Source Aggregation/Isolation

BGP announced prefix: **2001:db8::/32**

AS A — cybersecurity company

```
SOURCE IP
2001:db8:86e7:3637:106c:d7dc:e248:4a5d
2001:db8:2c7a:b1e7:e808:499c:d5b8:35b9
2001:db8:16cd:3fe3:3210:e49f:70f4:e081
2001:db8:3af5:a3e0:d5f1:8885:f3f3:da78
2001:db8:bd8:72c4:5b7e:01da7:88cc:99e1
2001:db8:69eb:ade2:a2f8:da13:11ed:5702
2001:db8:f1c5:3a12:3506:37eb:61c6:9322
2001:db8:b794:67d9:ec6c:38d7:daa3:71e9
2001:db8:a1f4:2409:f182:02d2:96c3:f96f
2001:db8:748e:22f1:fba1:0062:e3c6:8183
```

one single
scan entity
entire /32 prefix

Key Challenge: Source Aggregation/Isolation

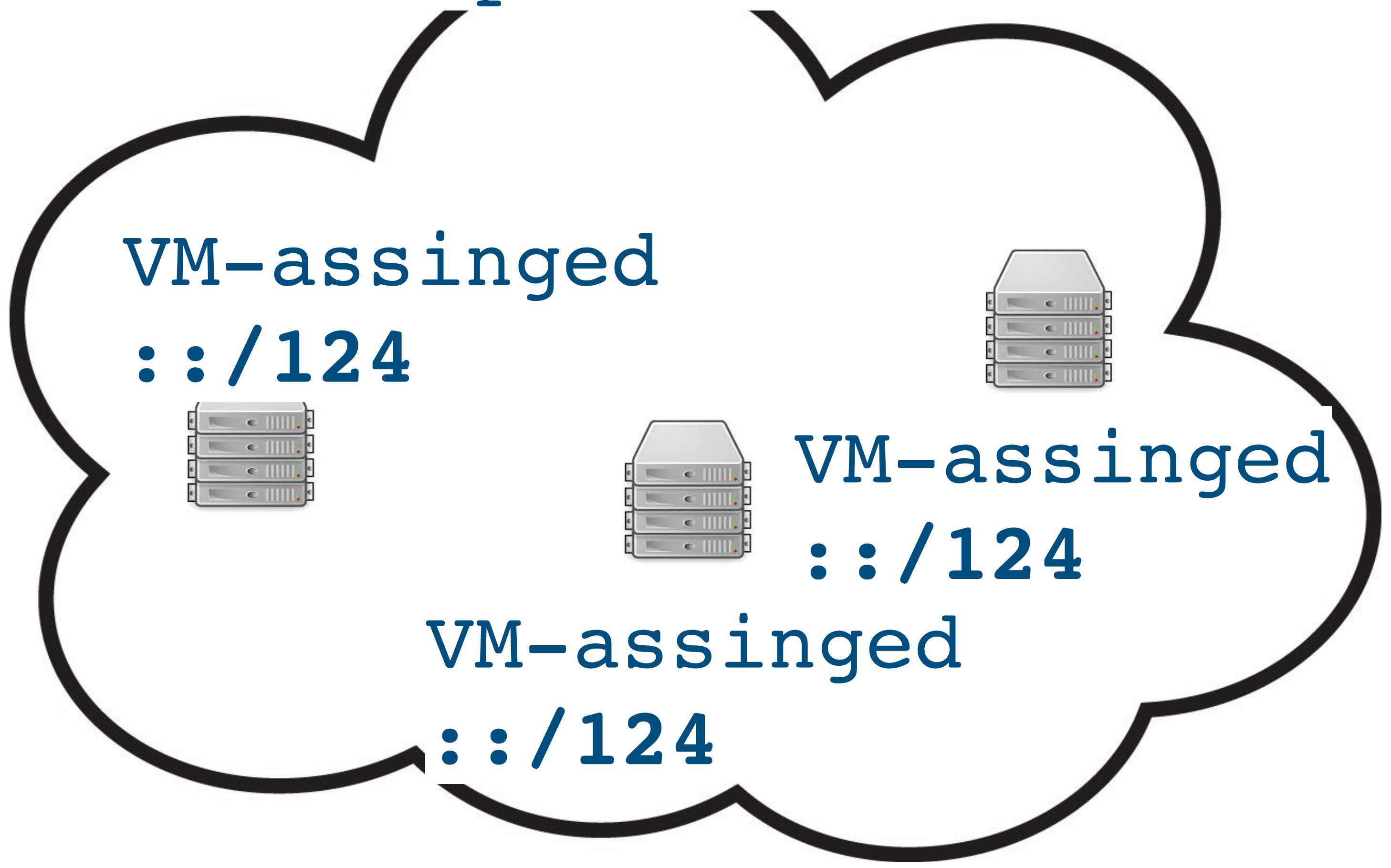
BGP announced prefix: **2001:db8::/32**

AS A — cybersecurity company

SOURCE IP
2001:db8:86e7:3637:106c:d7dc:e248:4a5d
2001:db8:2c7a:b1e7:e808:499c:d5b8:35b9
2001:db8:16cd:3fe3:3210:e49f:70f4:e081
2001:db8:3af5:a3e0:d5f1:8885:f3f3:da78
2001:db8:bd8:72c4:5b7e:01da7:88cc:99e1
2001:db8:69eb:ade2:a2f8:da13:11ed:5702
2001:db8:f1c5:3a12:3506:37eb:61c6:9322
2001:db8:b794:67d9:ec6c:38d7:daa3:71e9
2001:db8:a1f4:2409:f182:02d2:96c3:f96f
2001:db8:748e:22f1:fba1:0062:e3c6:8183

one single
scan entity
entire /32 prefix

BGP announced prefix: **2001:db9::/32**



AS B — major cloud provider

SOURCE IP
2001:db9:2143:11e4:6083:4e9f:aa01
2001:db9:2143:11e4:6083:4e9f:aa01
2001:db9:2143:11e4:6083:4e9f:aa01

scanner A
/124 prefix

2001:db9:2143:11e4:6083:4e9f:ba01
2001:db9:2143:11e4:6083:4e9f:ba01
2001:db9:2143:11e4:6083:4e9f:ba01

scanner B
/124 prefix

2001:db9:2143:11e4:6083:4e9f:ca01
2001:db9:2143:11e4:6083:4e9f:ca01
2001:db9:2143:11e4:6083:4e9f:ca01

scanner C
/124 prefix

Key Challenge: Source Aggregation/Isolation

AS A — cybersecurity company

```
SOURCE IP
2001:db8:86e7:3637:106c:d7dc:e248:4a5d
2001:db8:2c7a:b1e7:e808:499c:d5b8:35b9
2001:db8:16cd:3fe3:3210:e49f:70f4:e081
2001:db8:3af5:a3e0:d5f1:8885:f3f3:da78
2001:db8:bd8:72c4:5b7e:01da7:88cc:99e1
2001:db8:69eb:ade2:a2f8:da13:11ed:5702
2001:db8:f1c5:3a12:3506:37eb:61c6:9322
2001:db8:b794:67d9:ec6c:38d7:daa3:71e9
2001:db8:a1f4:2409:f182:02d2:96c3:f96f
2001:db8:748e:22f1:fba1:0062:e3c6:8183
```

one single
scan entity
entire /32 prefix

AS B — major cloud provider

```
SOURCE IP
2001:db9:2143:11e4:6083:4e9f:aa01
2001:db9:2143:11e4:6083:4e9f:aa01
2001:db9:2143:11e4:6083:4e9f:aa01
```

scanner A
/124 prefix

```
2001:db9:2143:11e4:6083:4e9f:ba01
2001:db9:2143:11e4:6083:4e9f:ba01
2001:db9:2143:11e4:6083:4e9f:ba01
```

scanner B
/124 prefix

```
2001:db9:2143:11e4:6083:4e9f:ca01
2001:db9:2143:11e4:6083:4e9f:ca01
2001:db9:2143:11e4:6083:4e9f:ca01
```

scanner C
/124 prefix

Without aggregation, we miss some (or all) of scanning activity!
With too much aggregation, we conflate scanners / block too much.

Key Findings

- The IPv6 space is actively being scanned!
- Detection - especially real-time - challenging
- More details in the paper!
    - Vantage points
    - Detection methodology
    - Details on services targeted, addresses targeted
    - And much more!

Illuminating Large-Scale IPv6 Scanning in the Internet

Philipp Richter
Akamai
prichter@akamai.com

Oliver Gasser
Max Planck Institute for Informatics
oliver.gasser@mpi-inf.mpg.de

Arthur Berger
Akamai/MIT
arthur@akamai.com

ABSTRACT

While scans of the IPv4 space are ubiquitous, today little is known about scanning activity in the IPv6 Internet. In this work, we present a longitudinal and detailed empirical study on large-scale IPv6 scanning behavior in the Internet, based on firewall logs captured at some 230,000 hosts of a major Content Distribution Network (CDN). We develop methods to identify IPv6 scans, assess current and past levels of IPv6 scanning activity, and study dominant characteristics of scans, including scanner origins, targeted services, and insights on how scanners find target IPv6 addresses. Where possible, we compare our findings to what can be assessed from publicly available traces. Our work identifies and highlights new challenges to detect scanning activity in the IPv6 Internet, and uncovers that today's scans of the IPv6 space show widely different characteristics when compared to the more well-known IPv4 scans.

CCS CONCEPTS

• **Networks** → **Network security**; **Network measurement**.

KEYWORDS

IPv6 scanning, Internet scanning, Internet security, network telescope, unsolicited traffic.

ACM Reference Format:
Philipp Richter, Oliver Gasser, and Arthur Berger. 2022. Illuminating Large-Scale IPv6 Scanning in the Internet. In *Proceedings of the 22nd ACM Internet Measurement Conference (IMC '22), October 25–27, 2022, Nice, France.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3517745.3561452

1 INTRODUCTION

Scanning the address space for vulnerable hosts and services is a key component in many of today's cyberattacks. In the IPv4 space, a scan of the entire address space can be conducted with comparably little resources in less than one hour [10], and botnets constantly scan the IPv4 space randomly to find new targets for infection [3]. This ubiquity of scanning activity in the IPv4 space makes scan detection readily possible, e.g., by leveraging darknets, or monitoring traffic on hosts or honeypots [22]. In the IPv6 Internet, both carrying out scans, as well as their detection, present a vastly more complicated task. Scanners can not simply target random addresses (there are more than 10^{38} IPv6 addresses) and must hence rely on hitlists or other heuristics to generate targets. At the same time, also the detection of IPv6 scans is challenging for two reasons: firstly, we need a vantage point that attracts and sees significant amounts of scanning traffic. Secondly, the vastness of the IPv6 space allows scanners to use entire subnets of varying sizes to emit scan traffic, potentially scanning from trillions of different source IP addresses, masking the true source of the scan traffic, and making scan detection difficult. Thus, conflating IPv6 and IPv4 scans, while tempting, presents a false equivalence. In this paper, we present a first-of-its-kind broad and longitudinal study of large-scale IPv6 scanning in the Internet. We make two key contributions:

Illuminating IPv6 scanning activity: We present detailed analyses on large-scale IPv6 scans carried out over the course of 15 months, as seen from a major CDN. We analyze scan sources, and study targeted services and addresses. We find that, unlike IPv4 scans, large-scale IPv6 scans are still comparably rare events, and we find them originating only from some 60 ASes. Further, IPv6 scan packets are concentrated on a small number of very active scan sources, with the two most active sources accounting for more than 70% of all logged scan traffic throughout our measurement window. Many large-scale IPv6 scans do not target a single or a small number of specific services, but rather scan large swaths of port numbers, sometimes exceeding 100 ports targeted per scan. This behavior more closely resembles general and unspecific penetration testing behavior, as opposed to scanning patterns of botnets trying to spread laterally by exploiting individual vulnerabilities. Our initial findings show that IPv6 scans in the wild show widely different characteristics from the more well-known IPv4 scans. We contrast our findings with what can be observed in publicly available data, and discuss potential reasons for our observations.

Measurement methodology: We identify key methodological challenges when it comes to pinpointing IPv6 scan sources and quantifying scanning activity and its properties. Regular IPv6 traffic is exchanged between two hosts using their 128-bit IPv6 addresses. However, in the case of scan traffic, we commonly find scanning actors not sourcing scan packets from an individual 128-bit source address, but from myriad source addresses spread across large prefixes. In such cases, any individual 128-bit source address used by a scanner may only emit very few packets (or even just a single packet), and thus hardly meet any criterion to be classified as a scan source. In fact, we find scanners using source addresses spread across prefixes as unspecific as a /32 prefix, a typical IPv6 allocation size for an *entire ISP*, thereby masking the true source of scanning activity. We show that when not aggregating source addresses to less-specific prefixes, such scanning activity may be missed in part or entirely, and can lead to severe misinterpretation of findings. Yet, in turn, too coarse aggregation of sources leads to conflating individual scan actors as well as non-scanning hosts. The methodological challenges faced in this work directly apply to scan detection and blocking in operational settings (e.g., Intrusion Detection Systems) and we argue that they present a looming major

IMC '22, October 25–27, 2022, Nice, France

ACM ISBN 978-1-4503-9259-4/22/10.
https://doi.org/10.1145/3517745.3561452