Scanning the Scanners

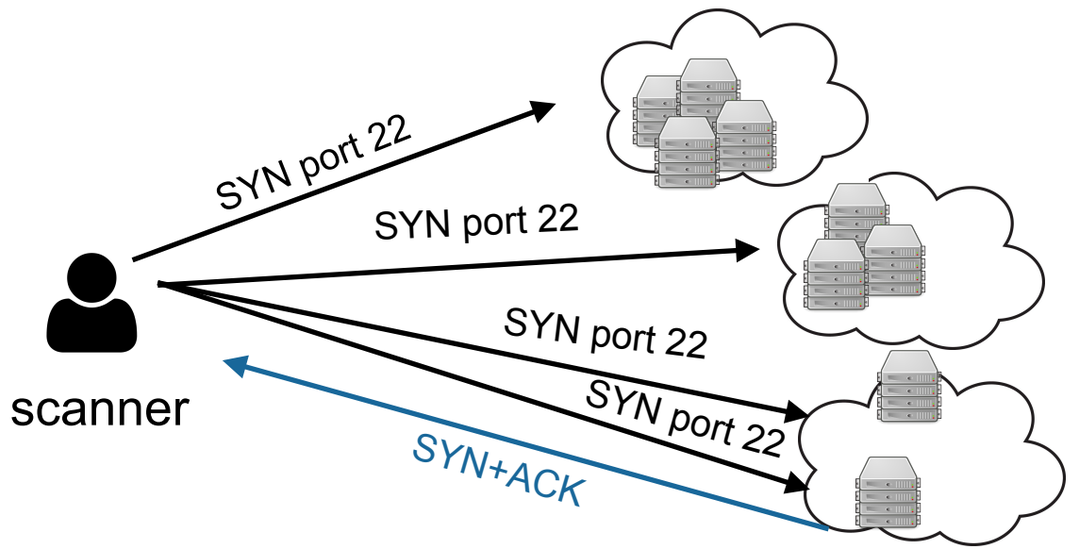
Sensing the Internet from a Massively Distributed Network Telescope

ACM Internet Measurement Conference 2019
Amsterdam, Netherlands

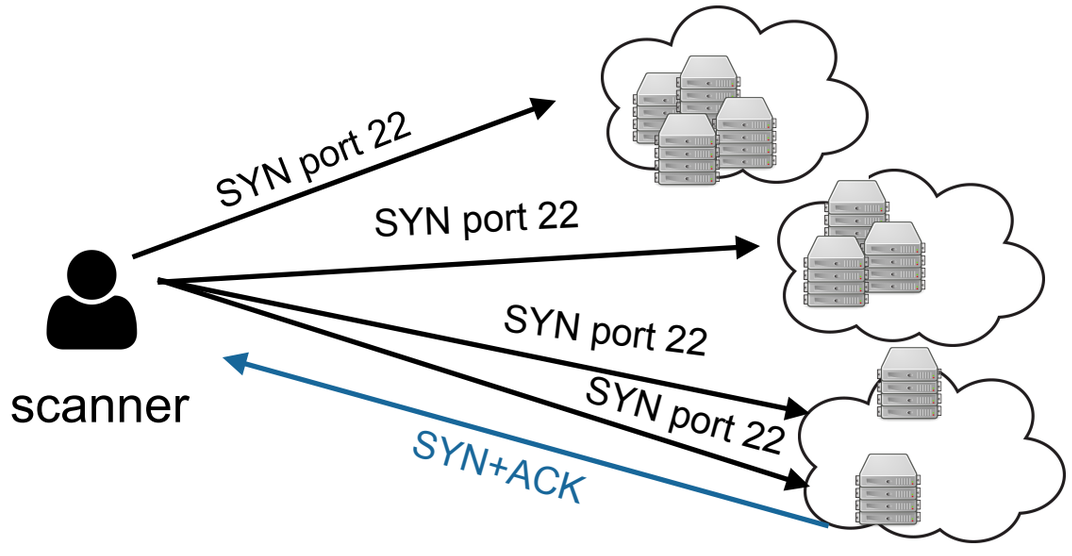
Philipp Richter and Arthur Berger

Massachusetts Institute of Technology
Akamai Technologies

Scanning 101

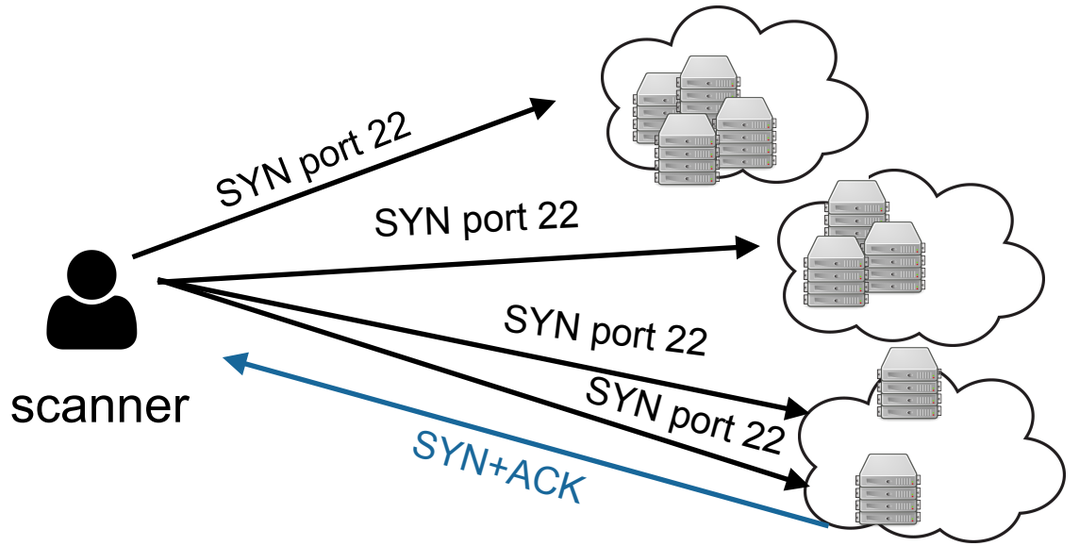


Scanning 101



1. Identify vulnerable services/devices
2. Exploit vulnerabilities!
   - Attack vulnerable hosts/networks directly *OR*
   - Form botnets to carry out cyberattacks against other targets

Scanning 101



1. Identify vulnerable services/devices
2. Exploit vulnerabilities!
   - Attack vulnerable hosts/networks directly *OR*
   - Form botnets to carry out cyberattacks against other targets

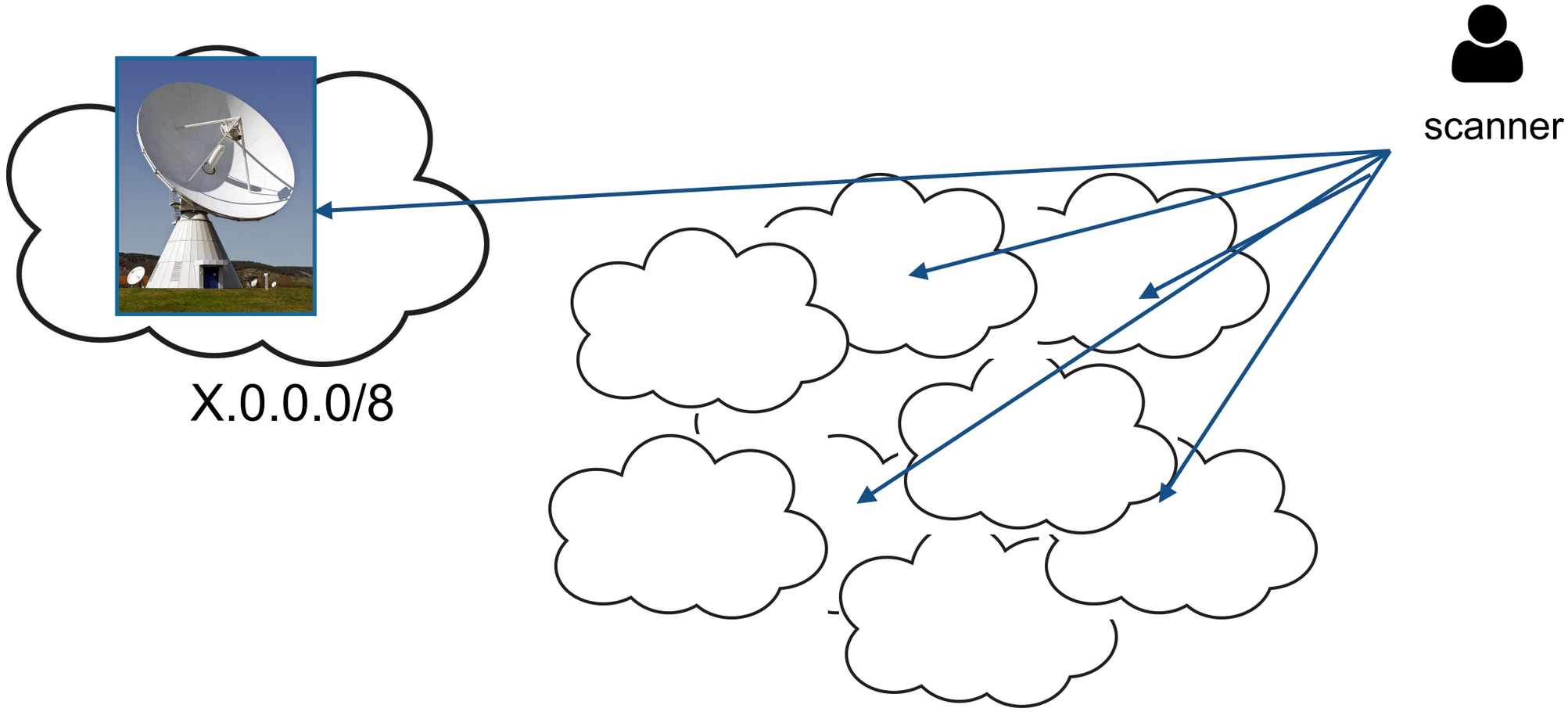
Scanning hosts on the Internet:
Key component in many of today's cyberattacks

Tracking Scanning Activity

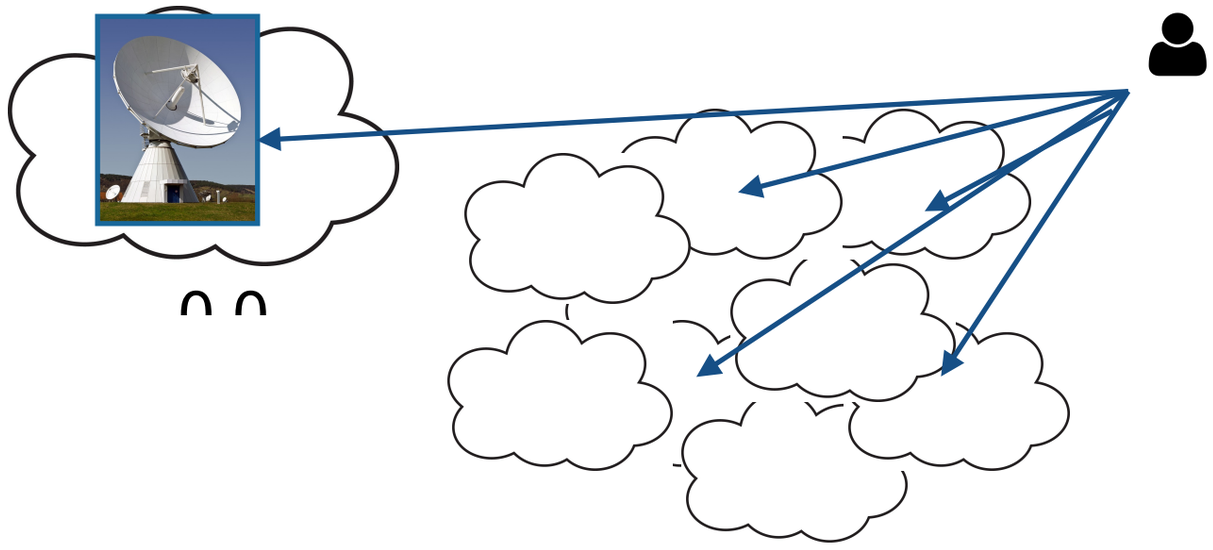
Tracking scanning activity: Assessment of ongoing and potentially upcoming cyberattacks

Detecting Scanning Activity: Darknets

Detecting Scanning Activity: Darknets



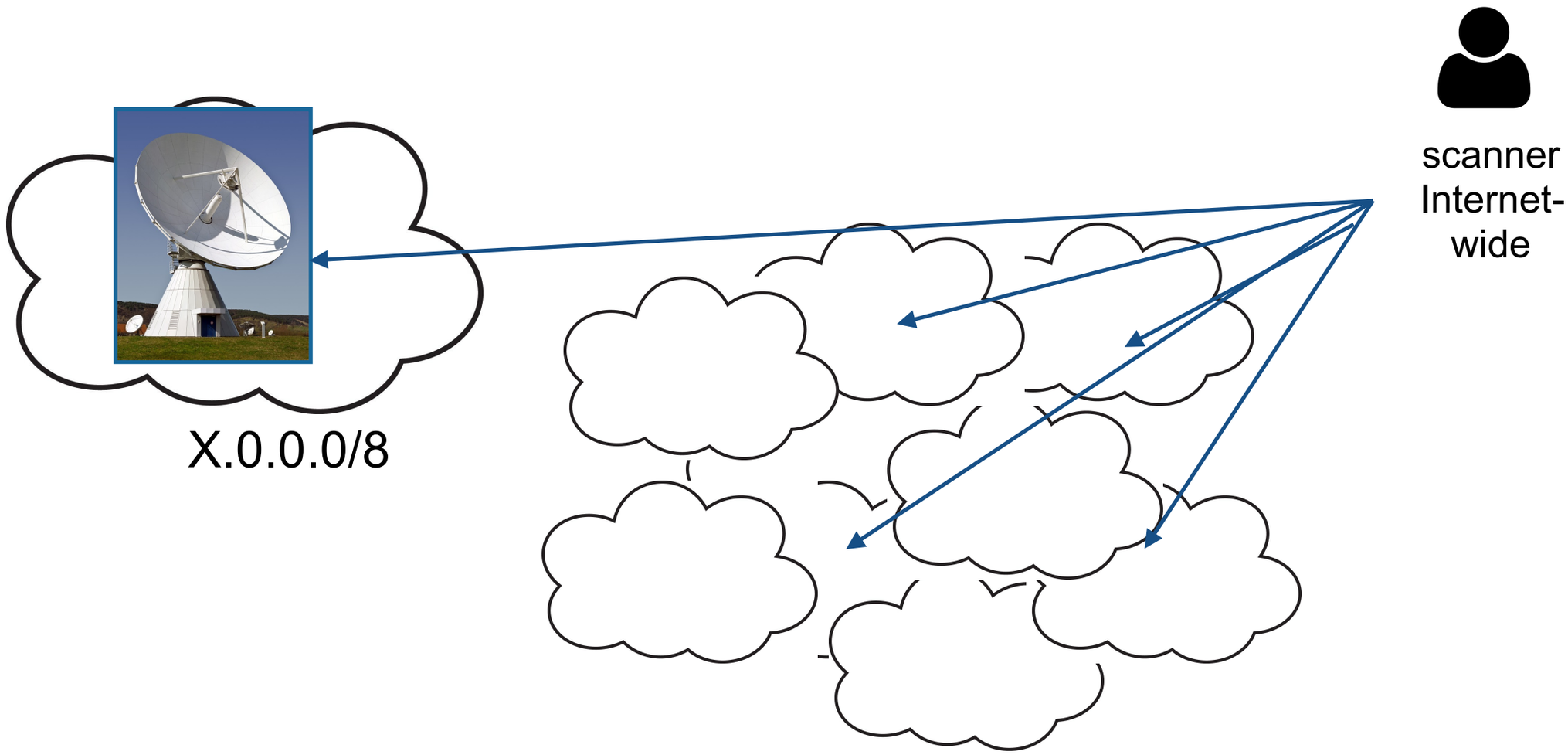
Internet-Wide Scans



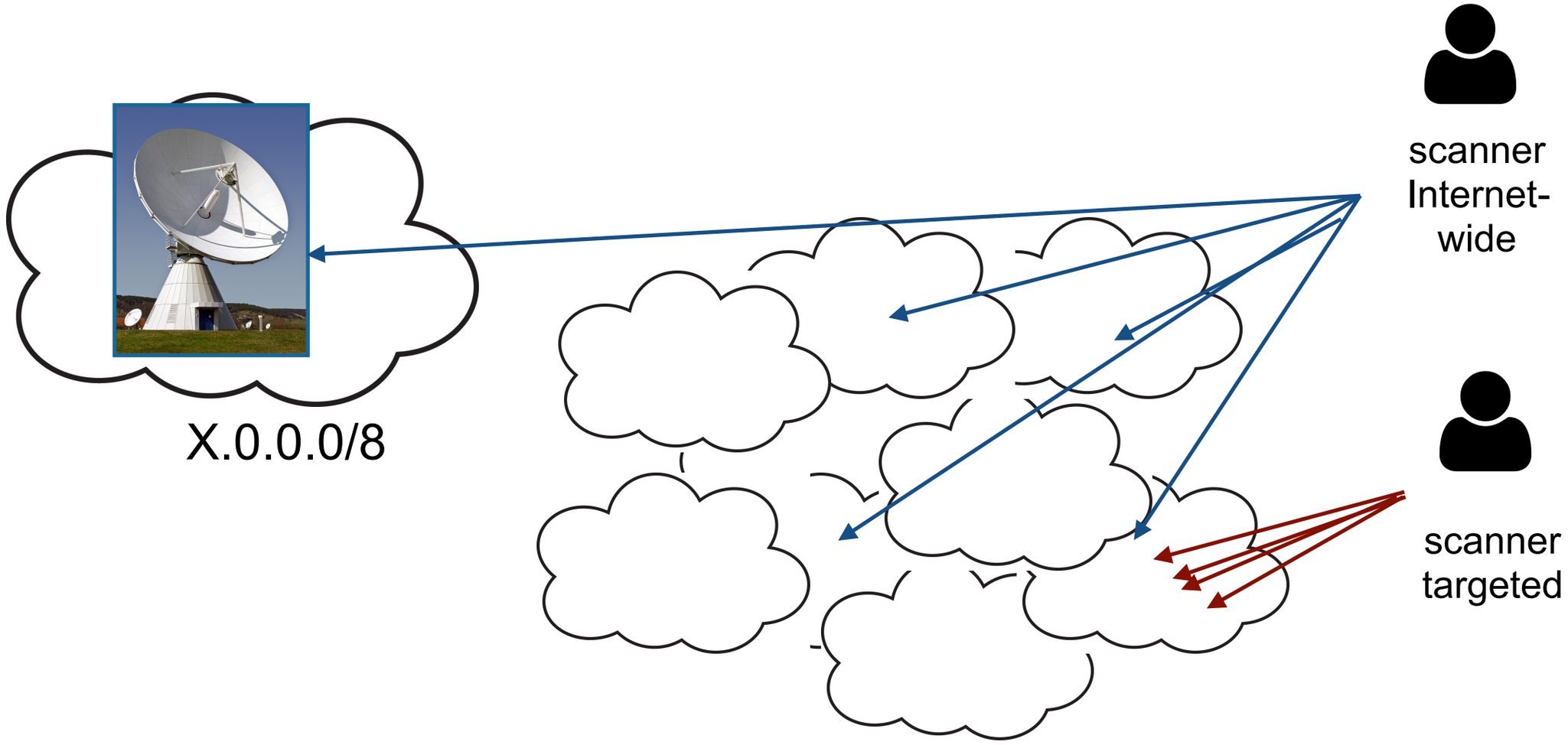
Darknets only capture “naive” Internet-wide scanning

- Scanner targets the entire IPv4 space (e.g., ZMap)
- Scanner targets a random subset of the IPv4 space

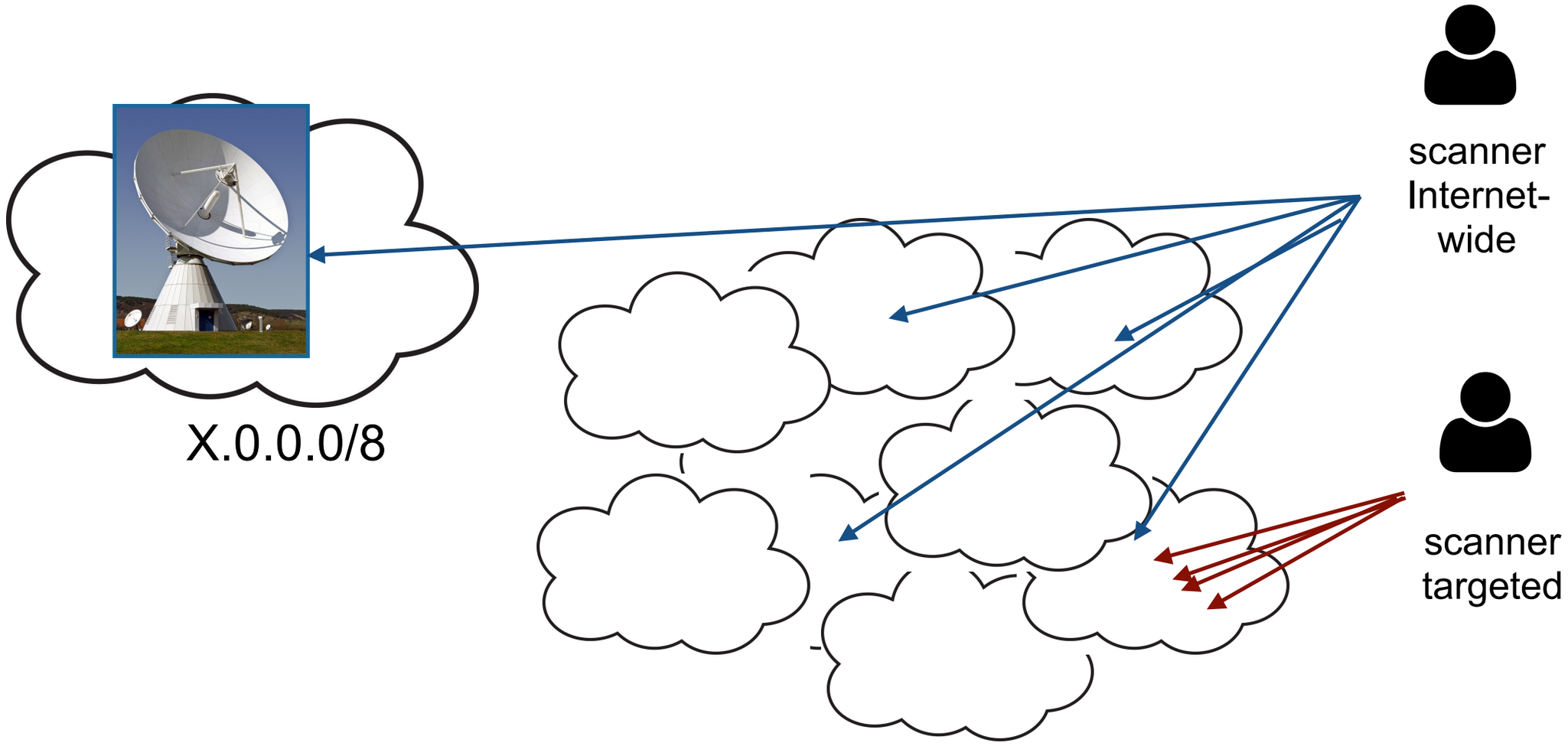
Internet-Wide vs. Targeted Scans



Internet-Wide vs. Targeted Scans



Internet-Wide vs. Targeted Scans



What about scans targeting specific prefixes or networks?

Undocumented and potentially much more dangerous!

Outline

Part 1

- What is a distributed network telescope?
- How much unsolicited traffic is there?

Part 2

- How much scanning traffic is there?
- How can we dissect scanning strategies?

Part 3

- How can an individual network detect targeted scans?

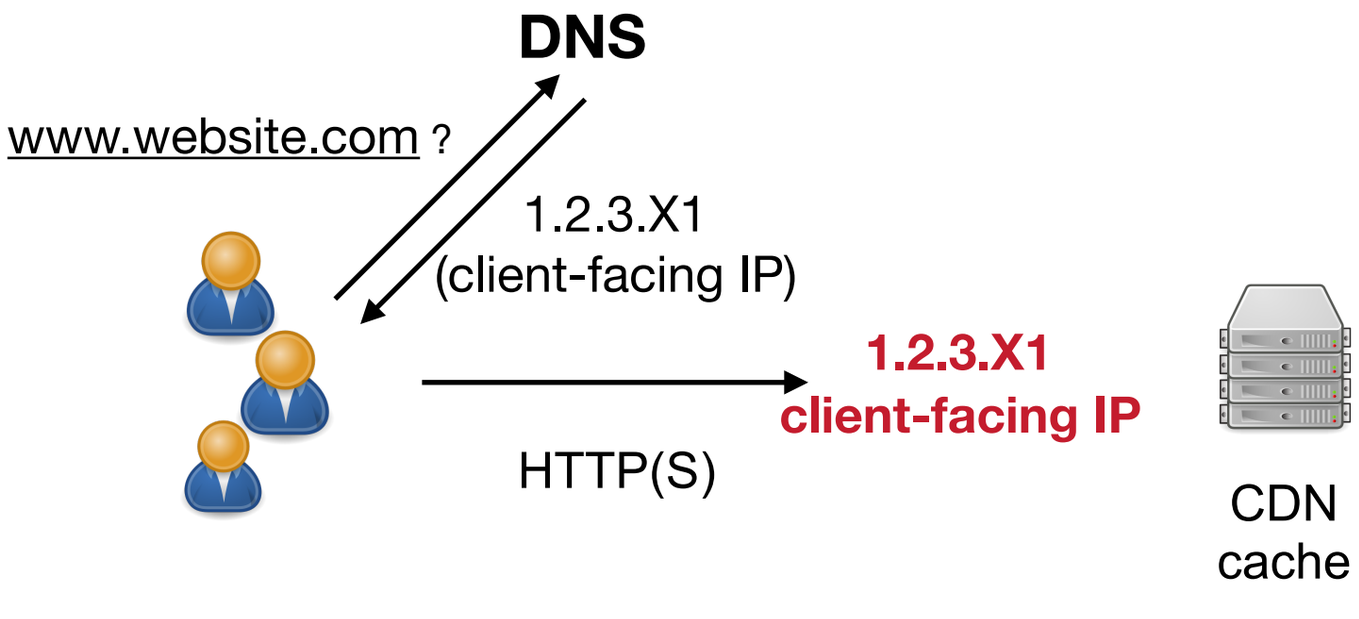
A Large CDN as a Network Telescope

- **89,000** CDN servers in **1,300+** ASes, **2,800+** BGP prefixes (a subset of the servers this CDN operates)

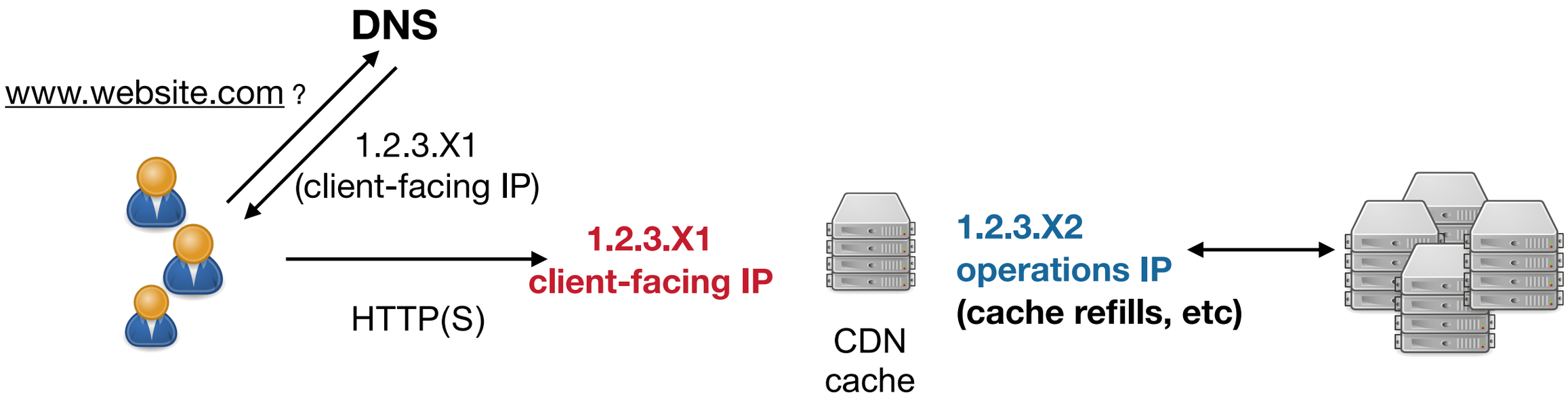
A Large CDN as a Network Telescope

- **89,000** CDN servers in **1,300+** ASes, **2,800+** BGP prefixes (a subset of the servers this CDN operates)
- Visibility into:
  - Internet-wide scans
  - Scans targeting the CDN itself
  - **Scans targeting networks that host CDN caches**

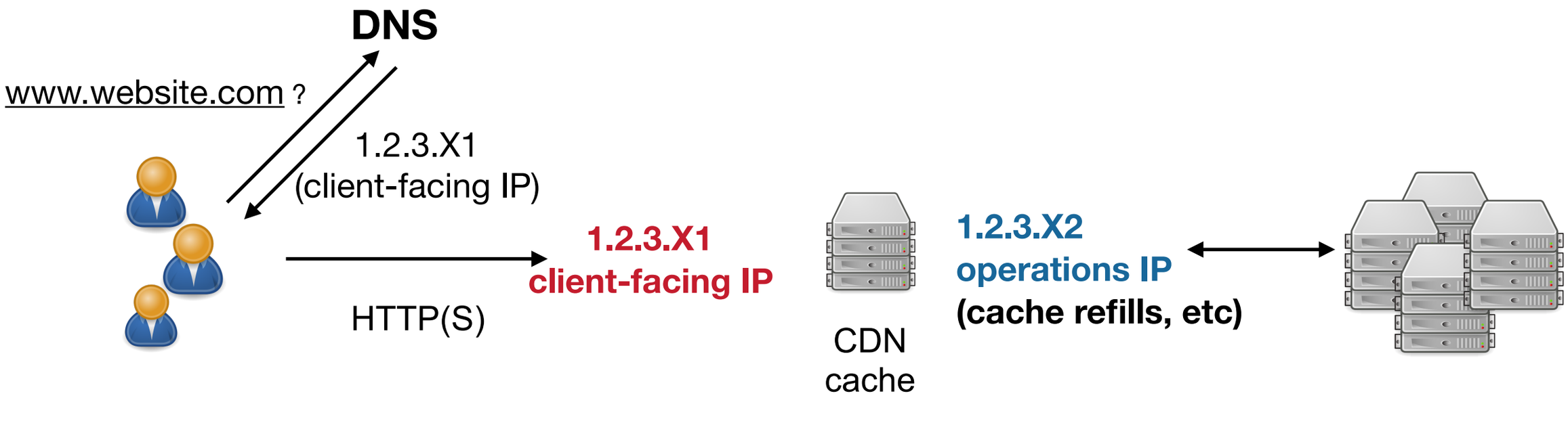
CDN Machine IPv4 Addresses



CDN Machine IPv4 Addresses



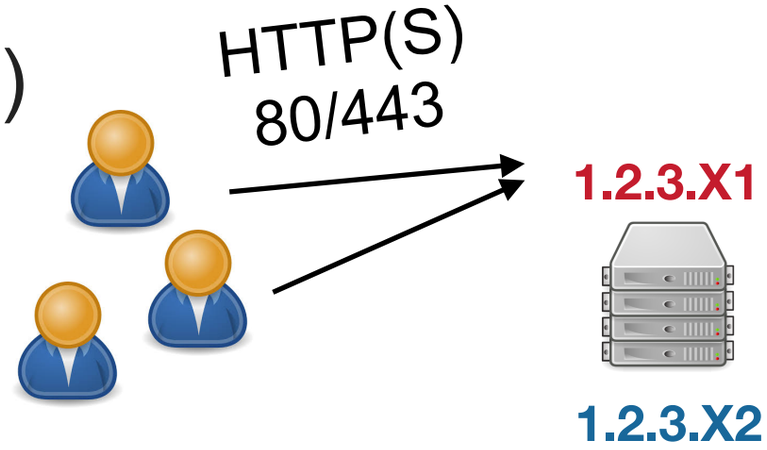
CDN Machine IPv4 Addresses



- Source IP hits only client-facing IPs of our servers:
  - CDN-targeted traffic
- Source IP hits client-facing and operations IPs 50/50:
  - CDN-agnostic traffic

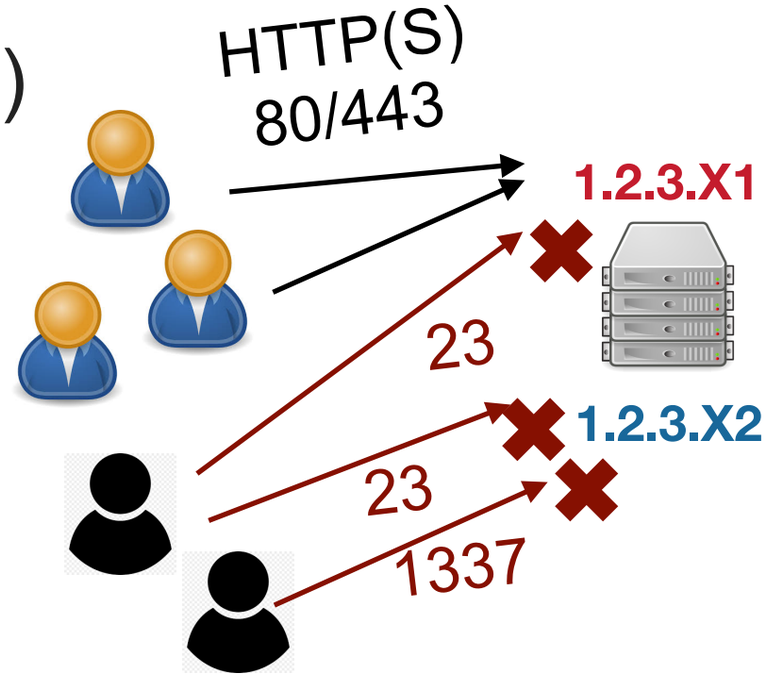
CDN Firewall Log Dataset

- Traffic to active port numbers (TCP 80/443,..)
  - Handled by web server/app. layer firewall
  - Not captured in our dataset



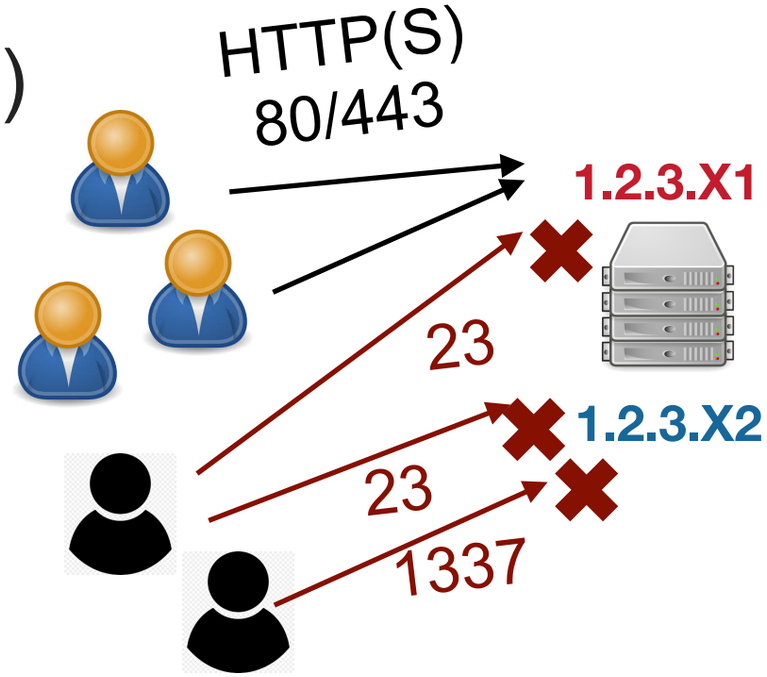
CDN Firewall Log Dataset

- Traffic to active port numbers (TCP 80/443,..)
  - Handled by web server/app. layer firewall
  - Not captured in our dataset
- Traffic to inactive port numbers (any IP)
  - Blocked by static firewall, logged
  - Too much traffic (e.g., DDoS) ⇨ “burst” ⇨ sampling triggered
  - 99.9% of time, a machine is NOT in “burst” state



CDN Firewall Log Dataset

- Traffic to active port numbers (TCP 80/443,..)
  - Handled by web server/app. layer firewall
  - Not captured in our dataset
- Traffic to inactive port numbers (any IP)
  - Blocked by static firewall, logged
  - Too much traffic (e.g., DDoS) ⇨ “burst” ⇨ sampling triggered
  - 99.9% of time, a machine is NOT in “burst” state

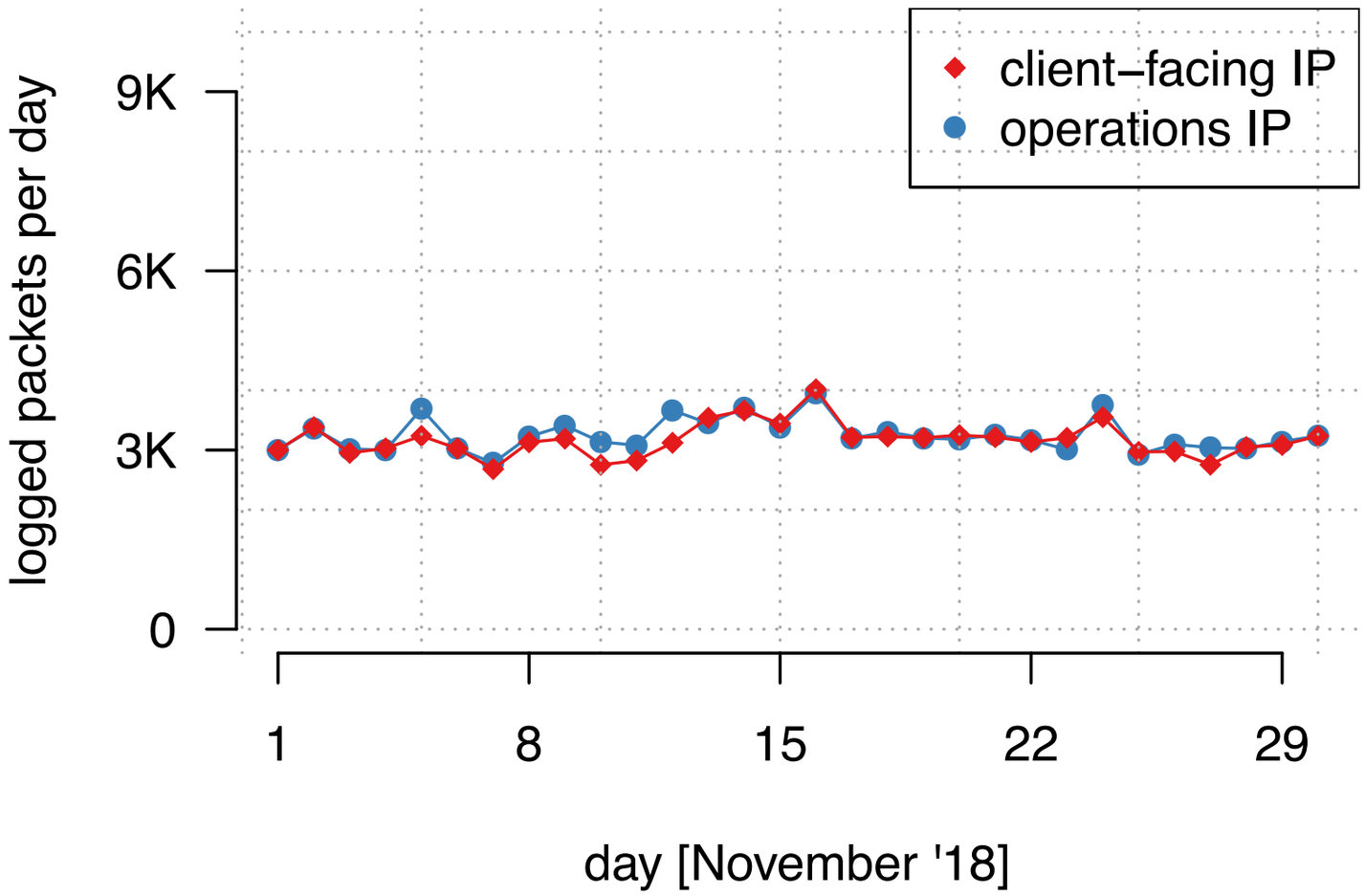


Full visibility of “sporadic” unsolicited traffic

total pkts	UDP	TCP	TCP SYN
19.4B	2.3B (12%)	17.1B (88%)	16.8B (86%)

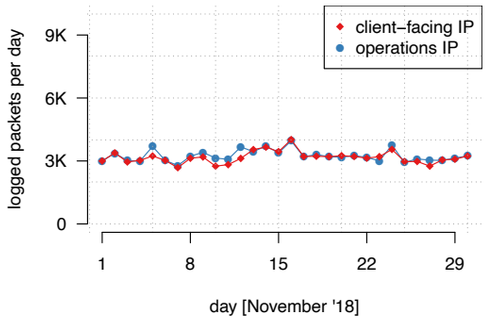
Dataset: November 2018

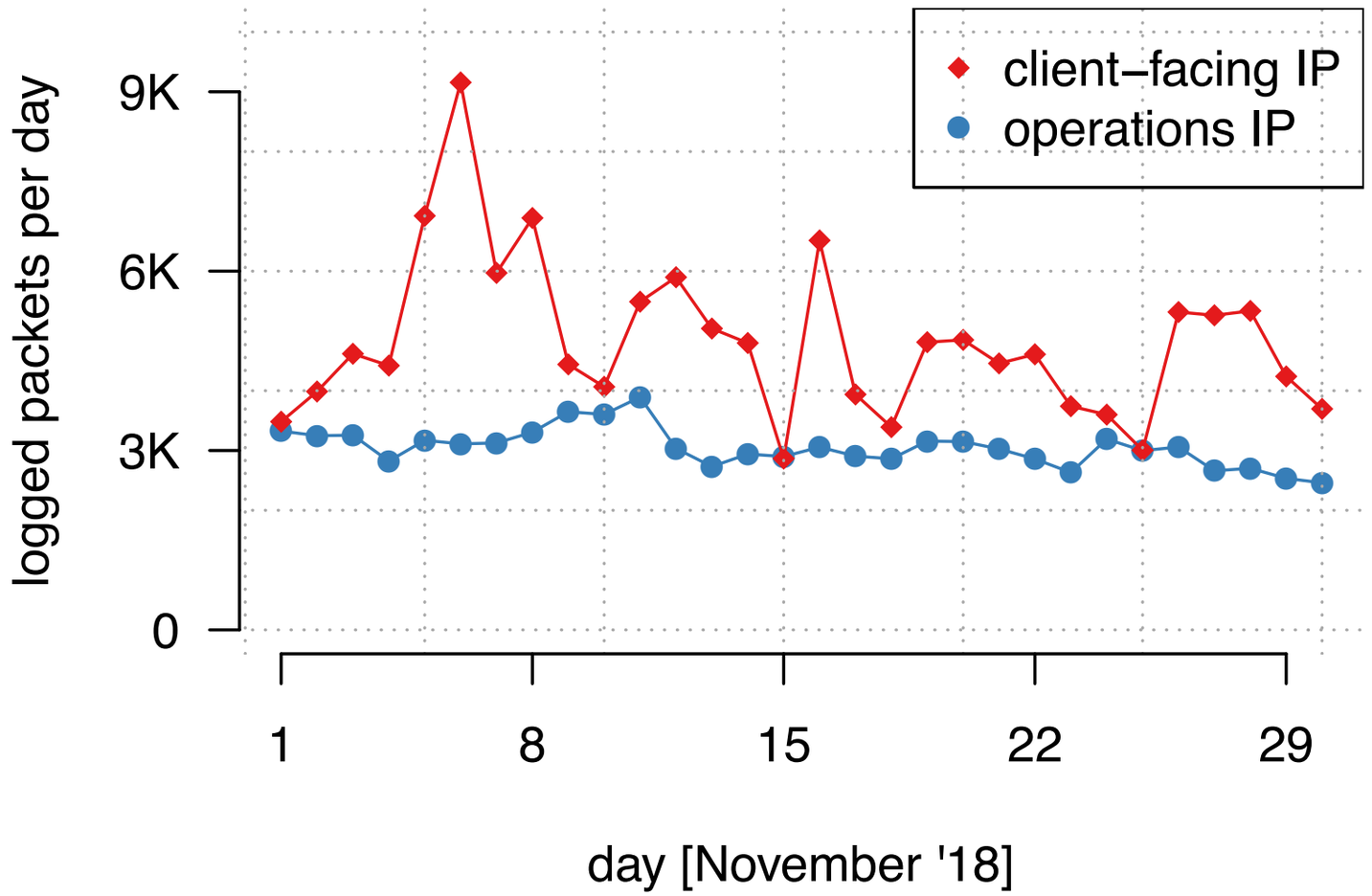
Unsolicited Traffic at CDN Caches



Machine 1: Baseline radiation

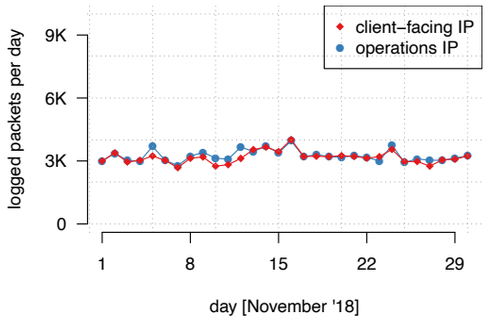
Unsolicited Traffic at CDN Caches

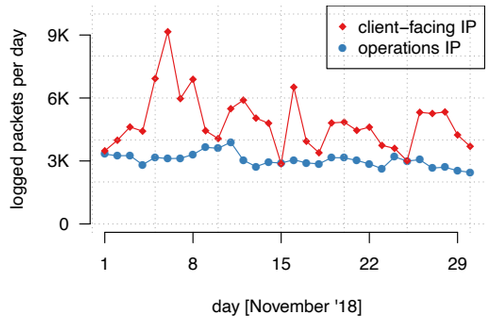


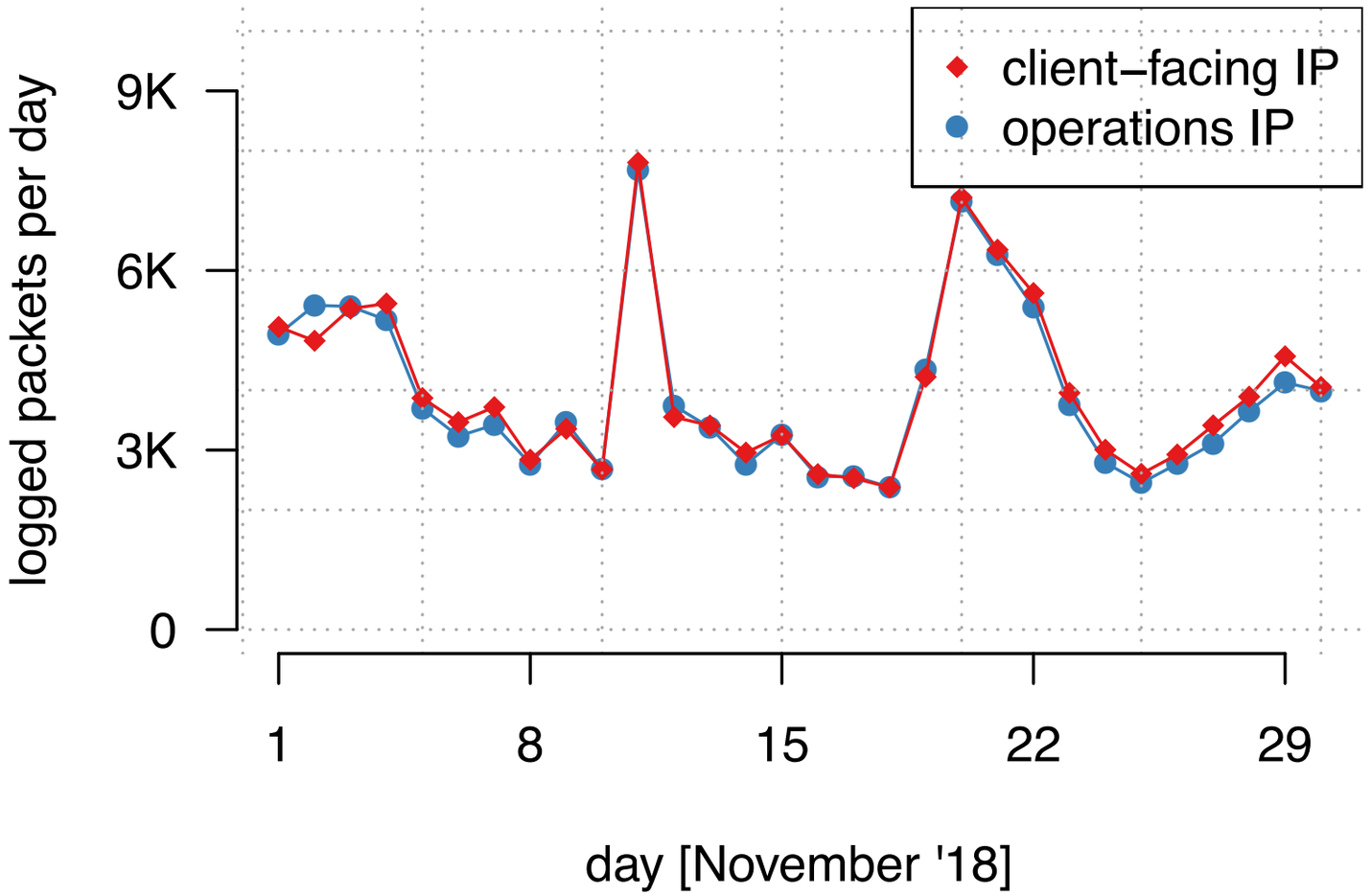


Machine 2: CDN-targeted spikes

Unsolicited Traffic at CDN Caches

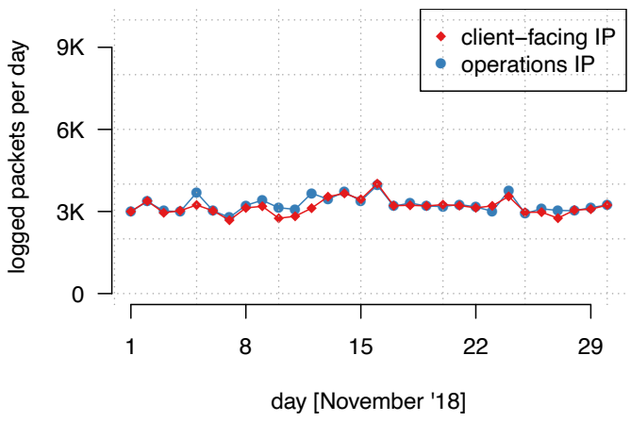


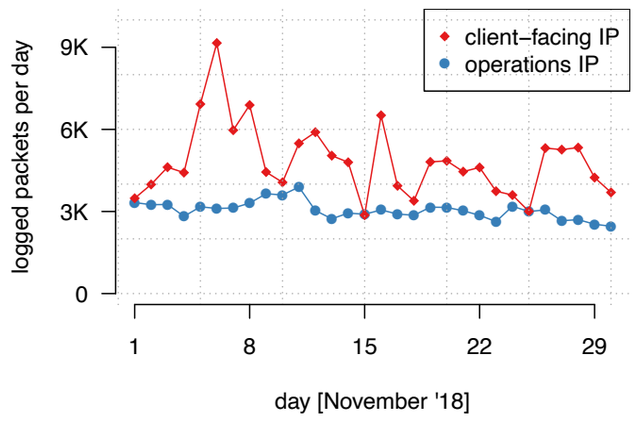


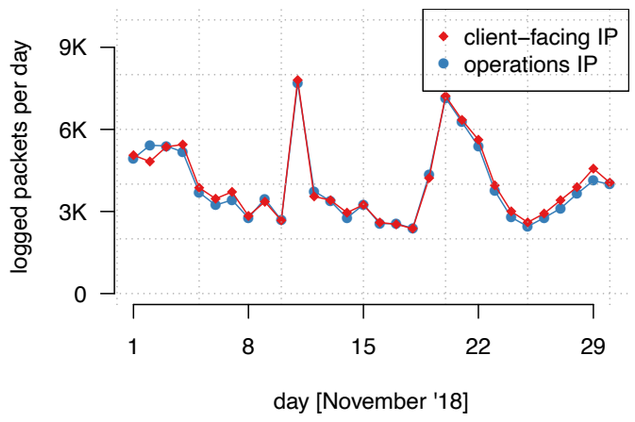


Machine 3: CDN-agnostic spikes

Unsolicited Traffic at CDN Caches

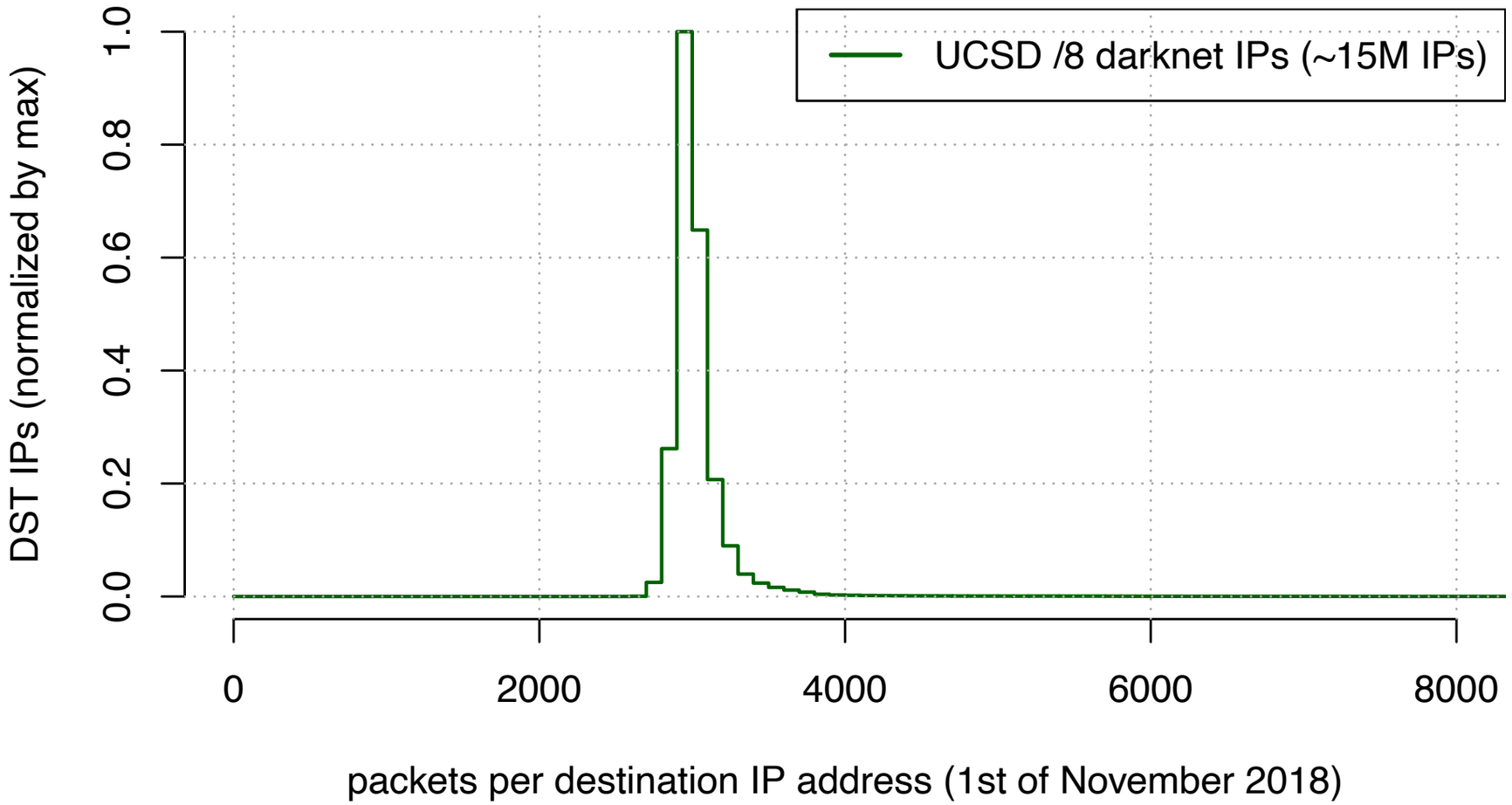






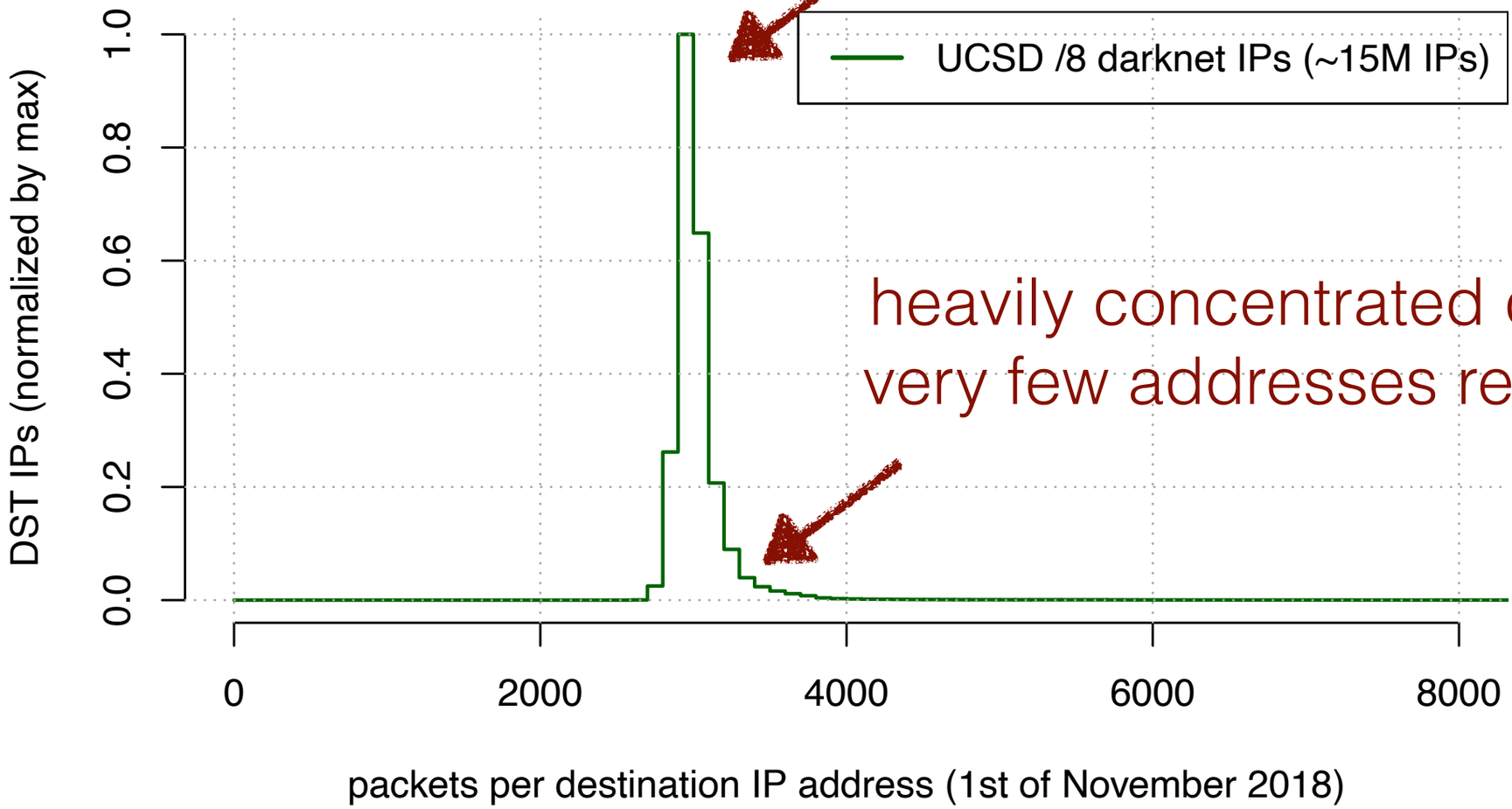
- Baseline radiation: **Every** IPv4 address: ~3000 pkts/day
- Spikes can be grouped into:
  - CDN-targeted ⇨ sources target domain names
  - CDN-agnostic ⇨ sources target prefixes/networks

Daily packets per DST IP in a /8 Darknet



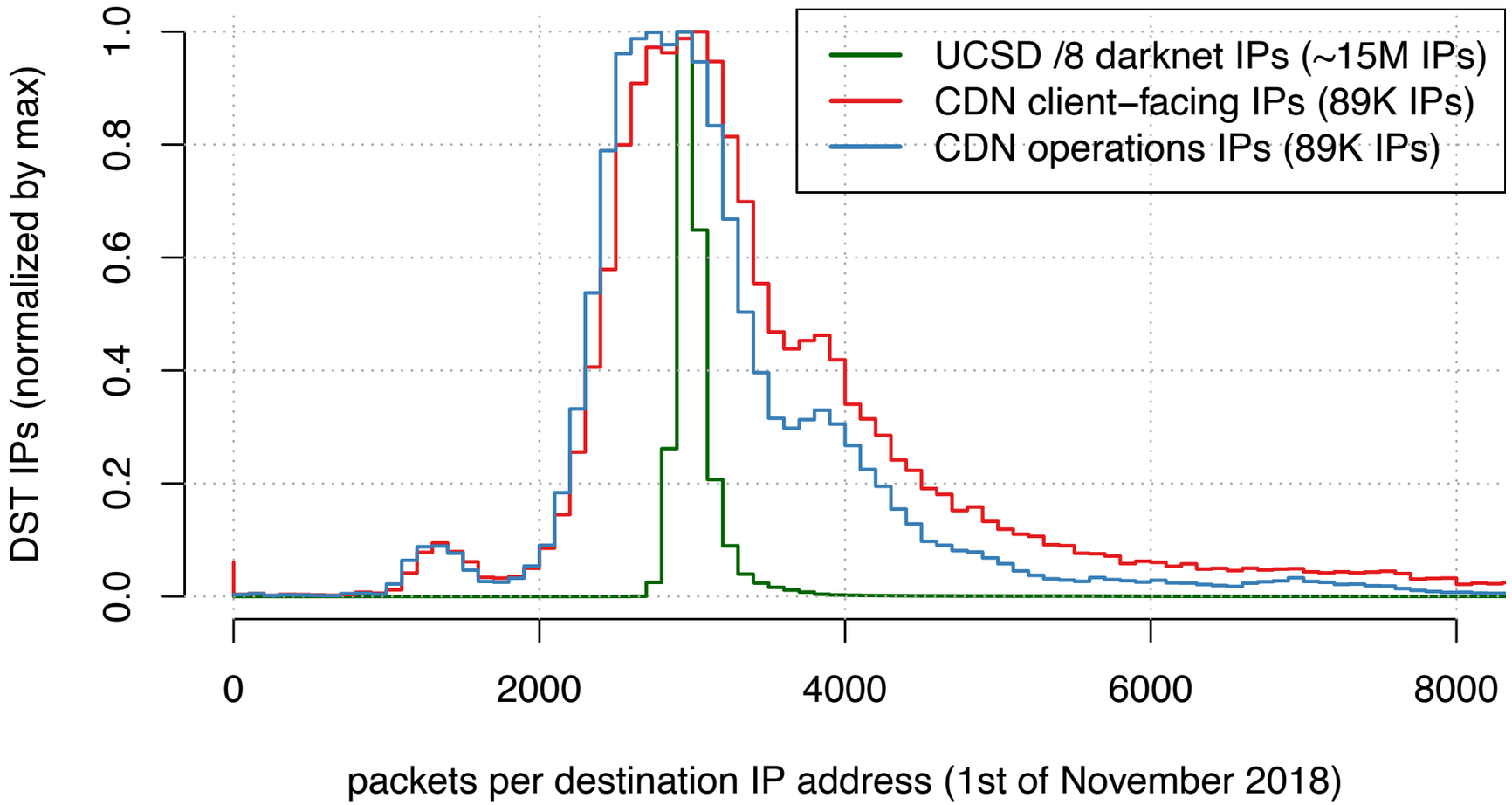
Daily packets per DST IP in a /8 Darknet

almost all darknet destination IP addresses:
baseline radiation - median: 3091 pkts/IP



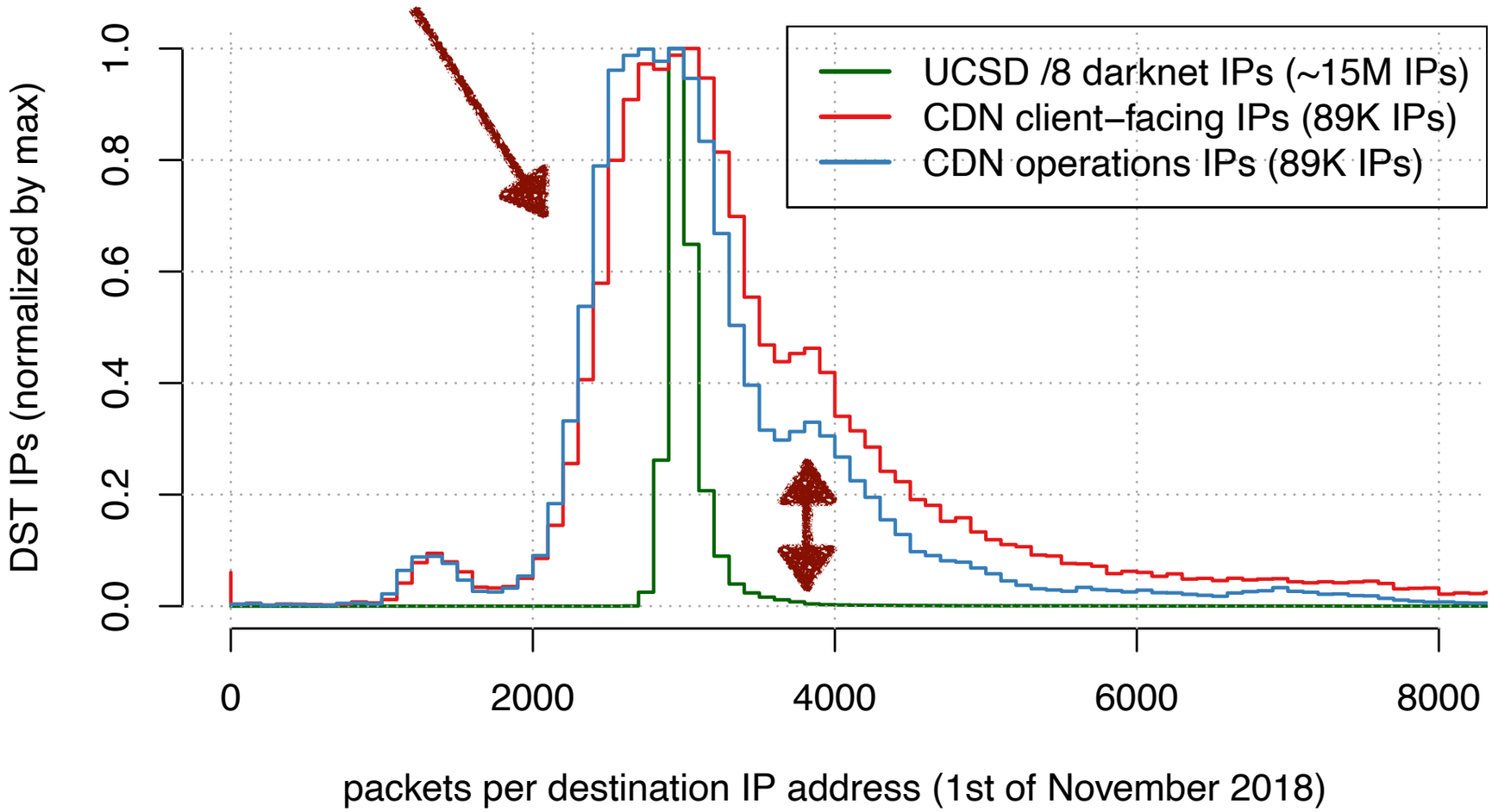
heavily concentrated distribution
very few addresses receive more

Daily packets per DST IP: Darknet vs. CDN

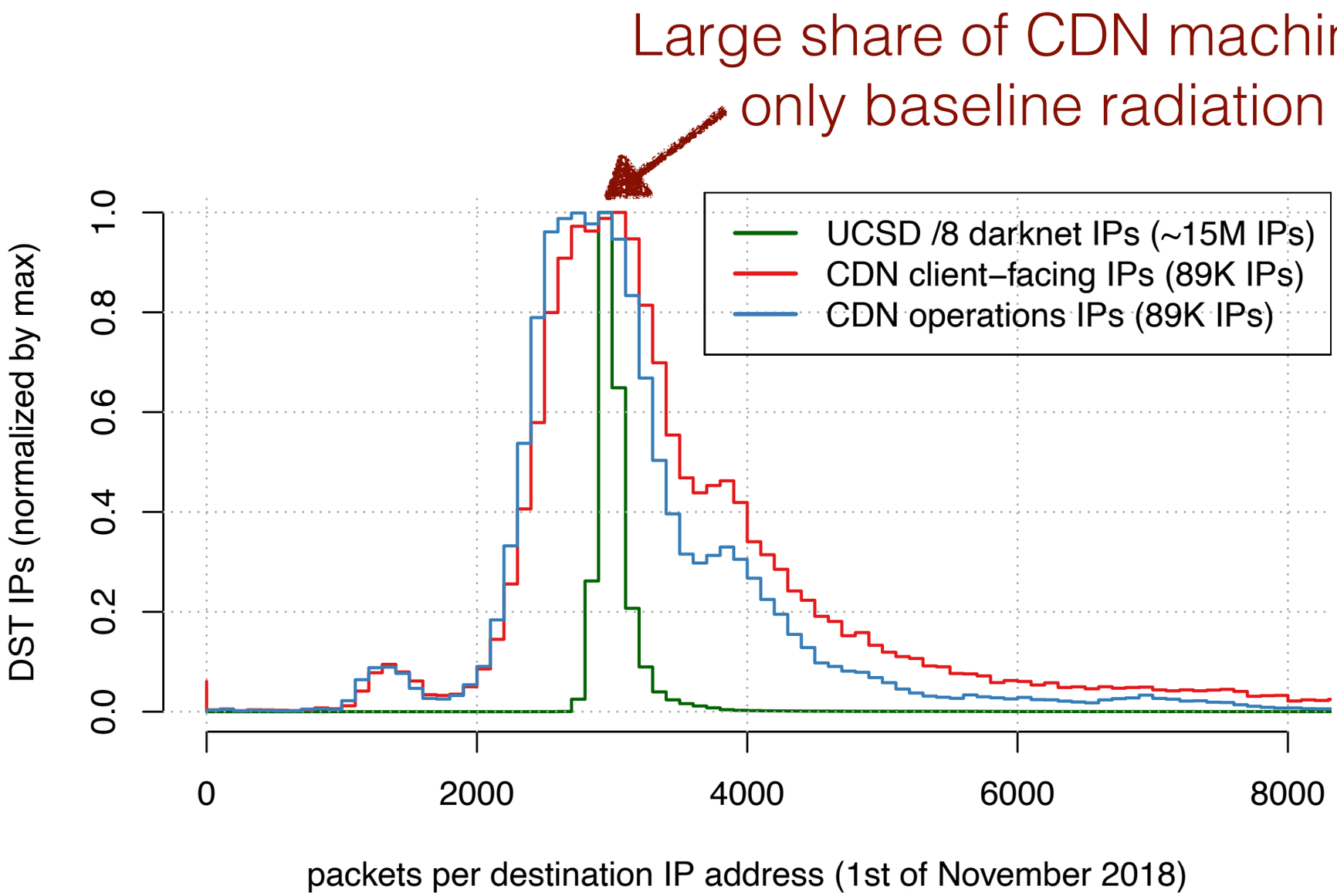


Daily packets per DST IP: Darknet vs. CDN

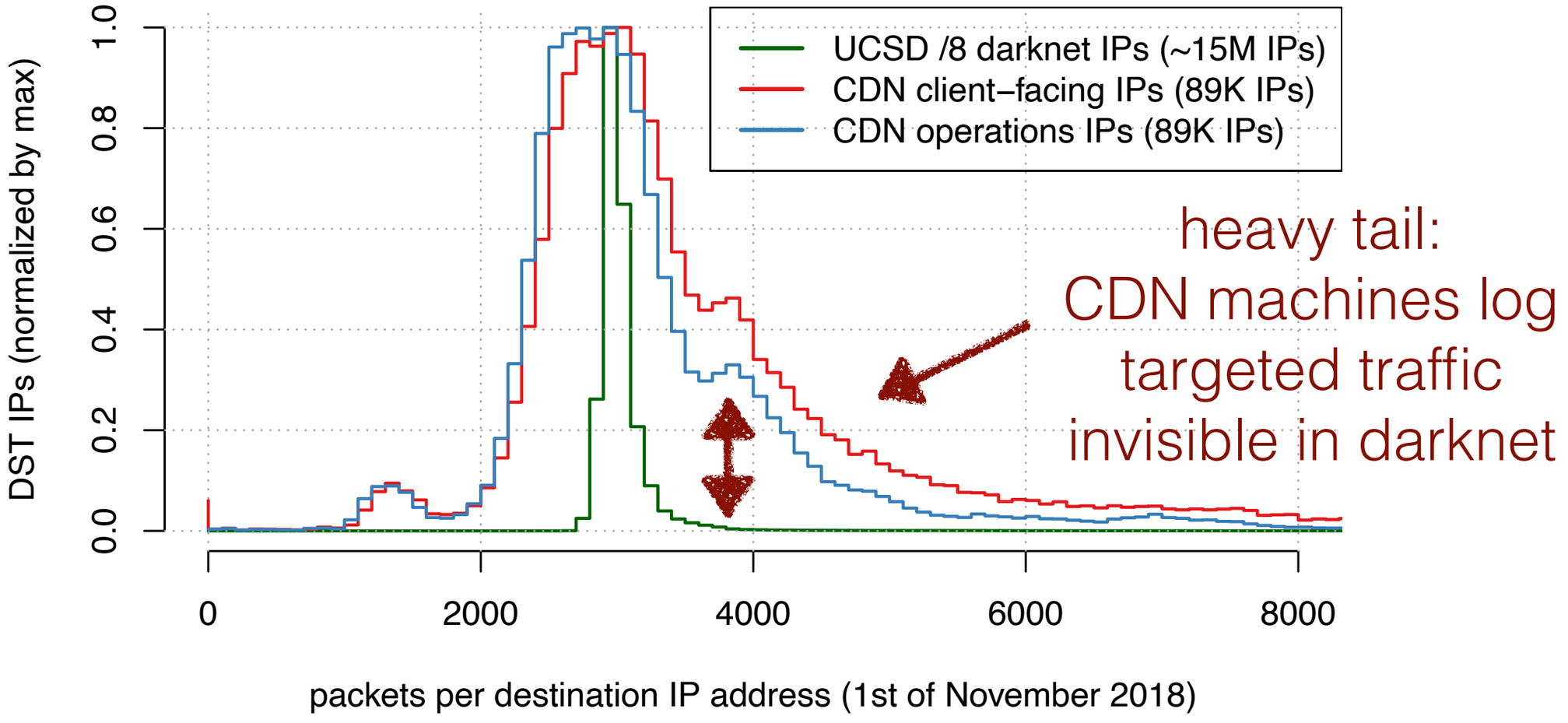
some CDN machines log
slightly less. Network filtering!



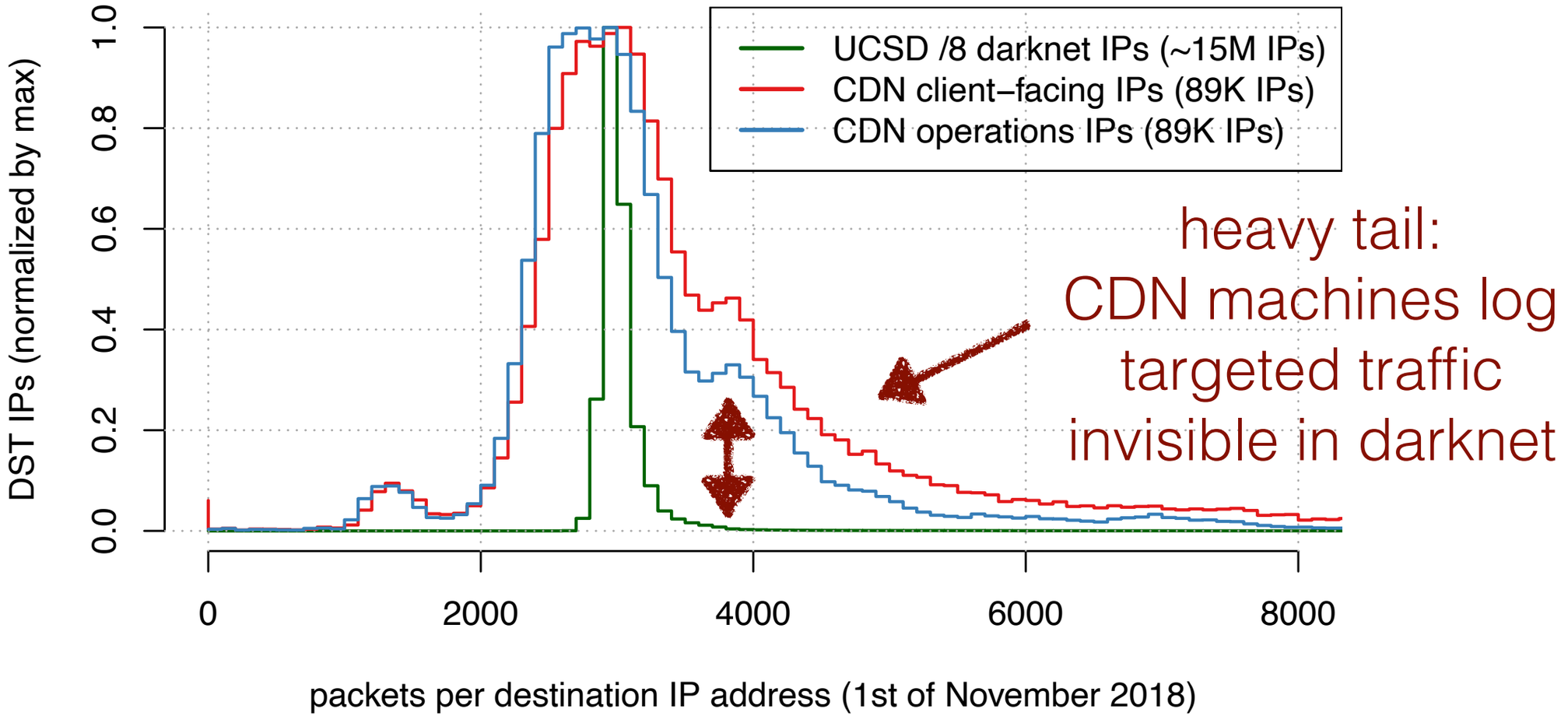
Daily packets per DST IP: Darknet vs. CDN



Daily packets per DST IP: Darknet vs. CDN



Daily packets per DST IP: Darknet vs. CDN



What is this additional unsolicited traffic?

Outline

Part 1

- What is a distributed network telescope?
- How much unsolicited traffic is there?

Part 2

- How much scanning traffic is there?
- How can we dissect scanning strategies?

Part 3

- How can an individual network detect targeted scans?

Identifying Scans

- Source IP contacts more than 100 destination IPs (IPv4)
- Maximum inter-arrival time 5400 seconds
- Effectively removes remaining DDoS attacks against CDN

Identifying Scans

- Source IP contacts more than 100 destination IPs (IPv4)
- Maximum inter-arrival time 5400 seconds
- Effectively removes remaining DDoS attacks against CDN

- **87%** of all logged unsolicited packets are part of scans
- November ’18: 2.17M scans, 16.8B packets, 1.1M SRC IPs

Scan Target Selection Strategies

all scans

Internet-wide

(source targets entire IPv4 space)

targeted

(source targets specific IPs/ranges)

Scan Target Selection Strategies

all scans

Internet-wide

(source targets entire IPv4 space)

targeted

(source targets specific IPs/ranges)

full IPv4 space

28% pkts

full IPv4 space: scanner IP targets
most/all of our IP addresses (e.g., ZMap)

Scan Target Selection Strategies

all scans

Internet-wide

(source targets entire IPv4 space)

targeted

(source targets specific IPs/ranges)

full IPv4 space

28% pkts

domain scans: targeted at DNS-exposed IPs

4.1% pkts

domain scans: scanner IP targets client-facing (DNS-exposed) IPs exclusively
E.g., scanning Alexa Top 1M etc.

Scan Target Selection Strategies

all scans

Internet-wide

(source targets entire IPv4 space)

full IPv4 space

28% pkts

targeted

(source targets specific IPs/ranges)

domain scans: targeted at DNS-exposed IPs

4.1% pkts

Leaves us with 2/3s of scan packets that do *not* target the full space *nor* target our CDN DNS-exposed IPs:

Scan Target Selection Strategies

all scans

Internet-wide

(source targets entire IPv4 space)

full IPv4 space

28% pkts

targeted

(source targets specific IPs/ranges)

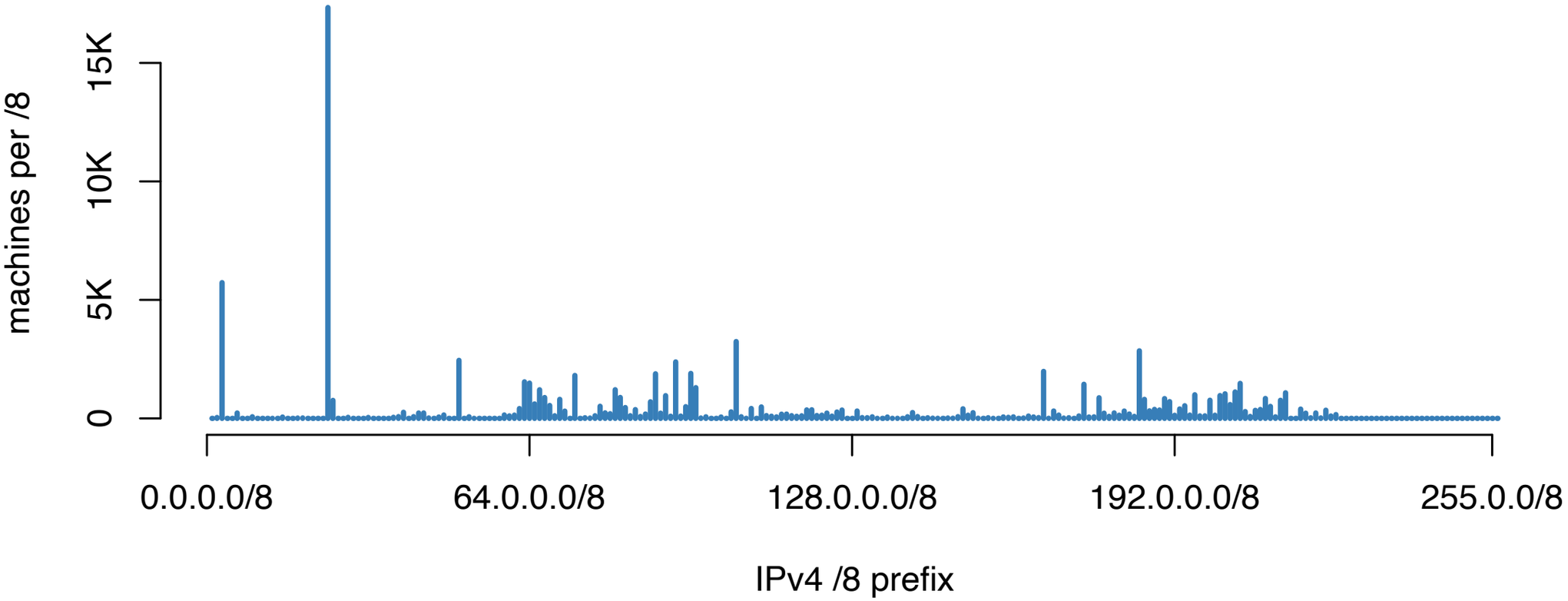
domain scans: targeted at DNS-exposed IPs

4.1% pkts

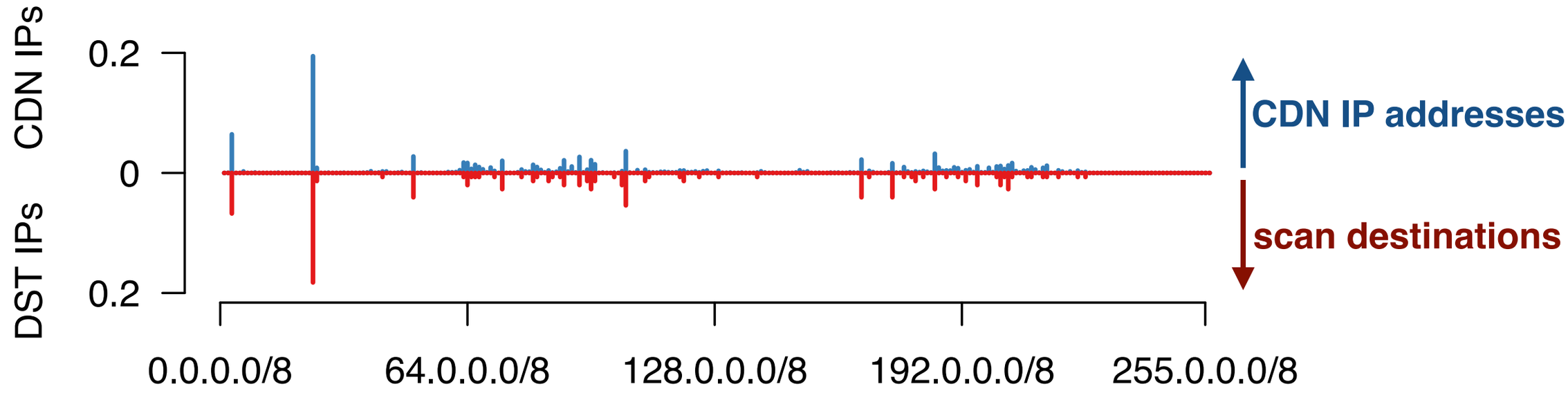
Leaves us with 2/3s of scan packets that do *not* target the full space *nor* target our CDN DNS-exposed IPs:

Do they target a random subset of the space or target specific prefixes/networks?

Detecting if Scan Targets Random Subset

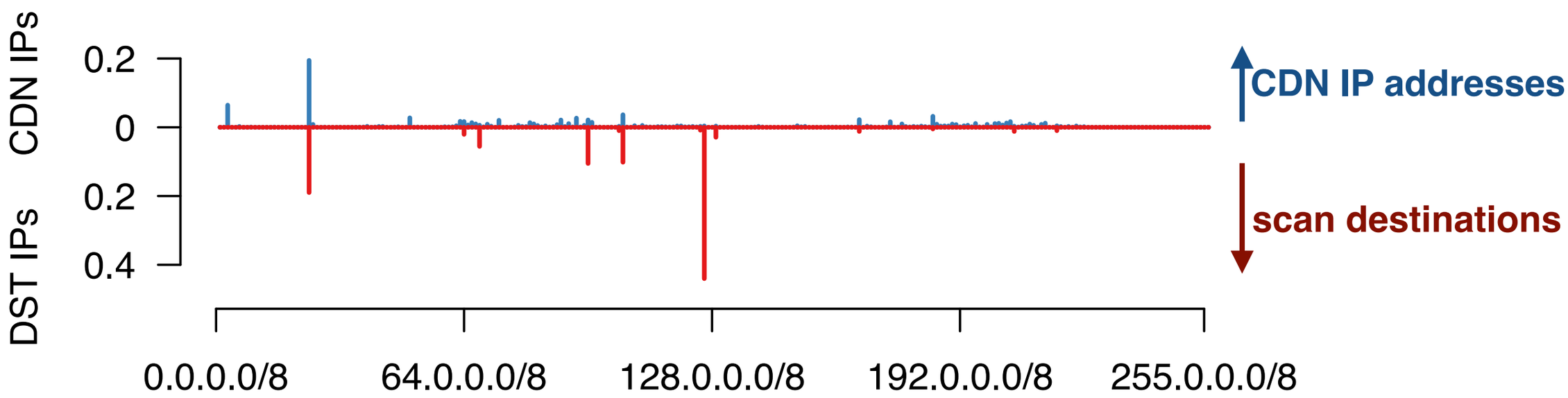


Internet-wide partial (random) scan



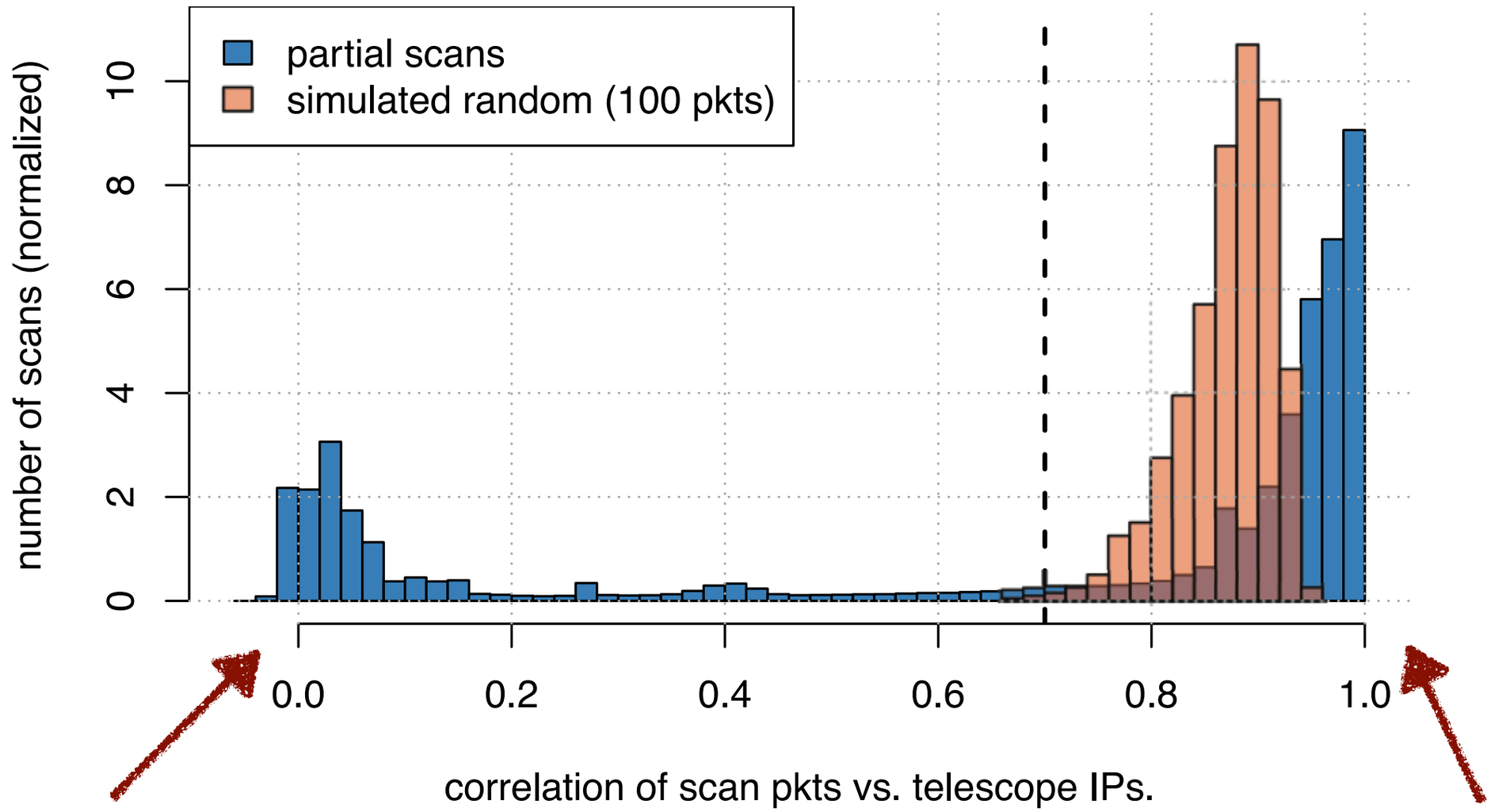
- Packets in scan total: *148*.
- Pearson correlation DST IPs / CDN IPs = *0.96.*
- This scan targets a random subset of CDN IPs.
  - **Internet-wide Partial scan.**

Localized Scan



- Packets in scan total: *827*.
- Pearson correlation DST IPs / CDN IPs = *0.38.*
- This scan targets a non-random subset of CDN IPs.
  - **Localized scan**

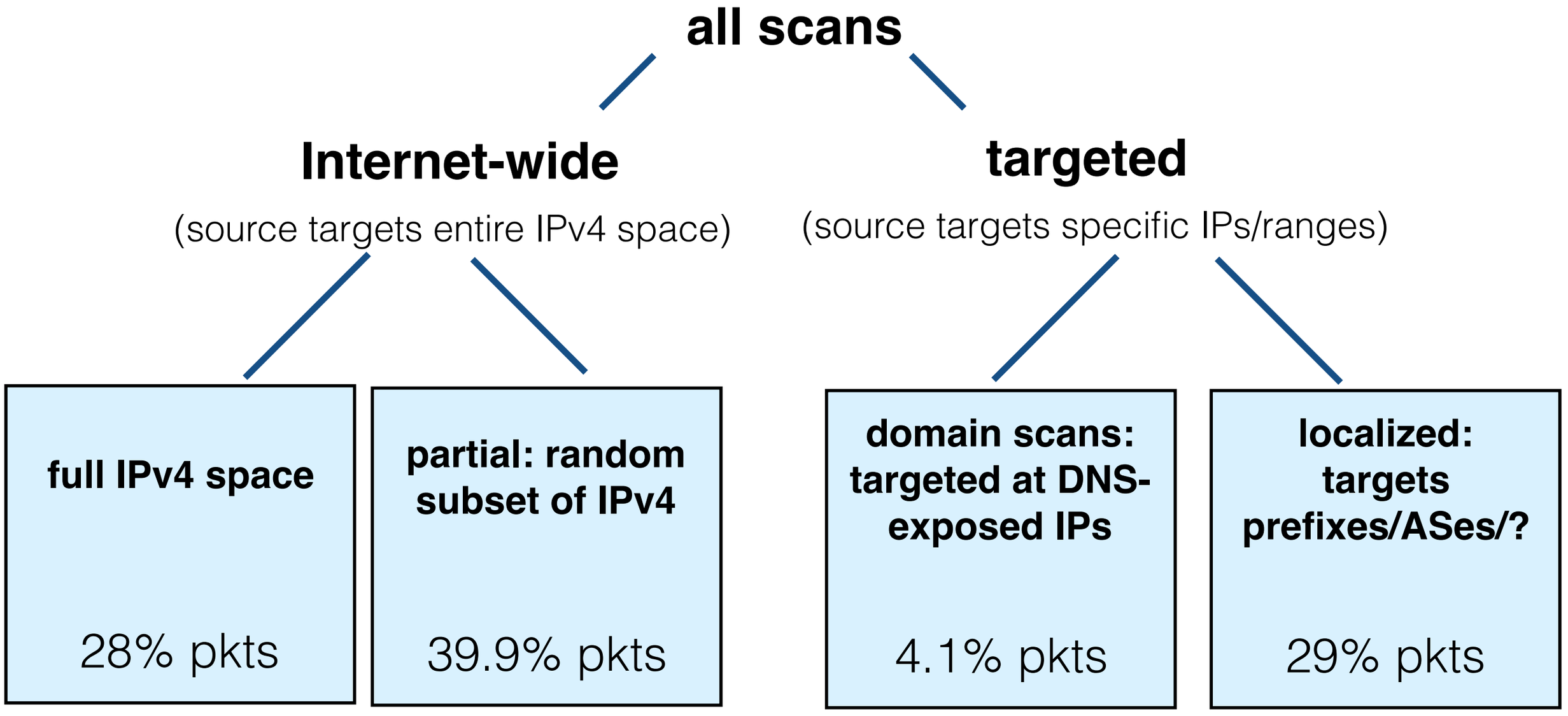
Internet-wide vs. Localized Scans



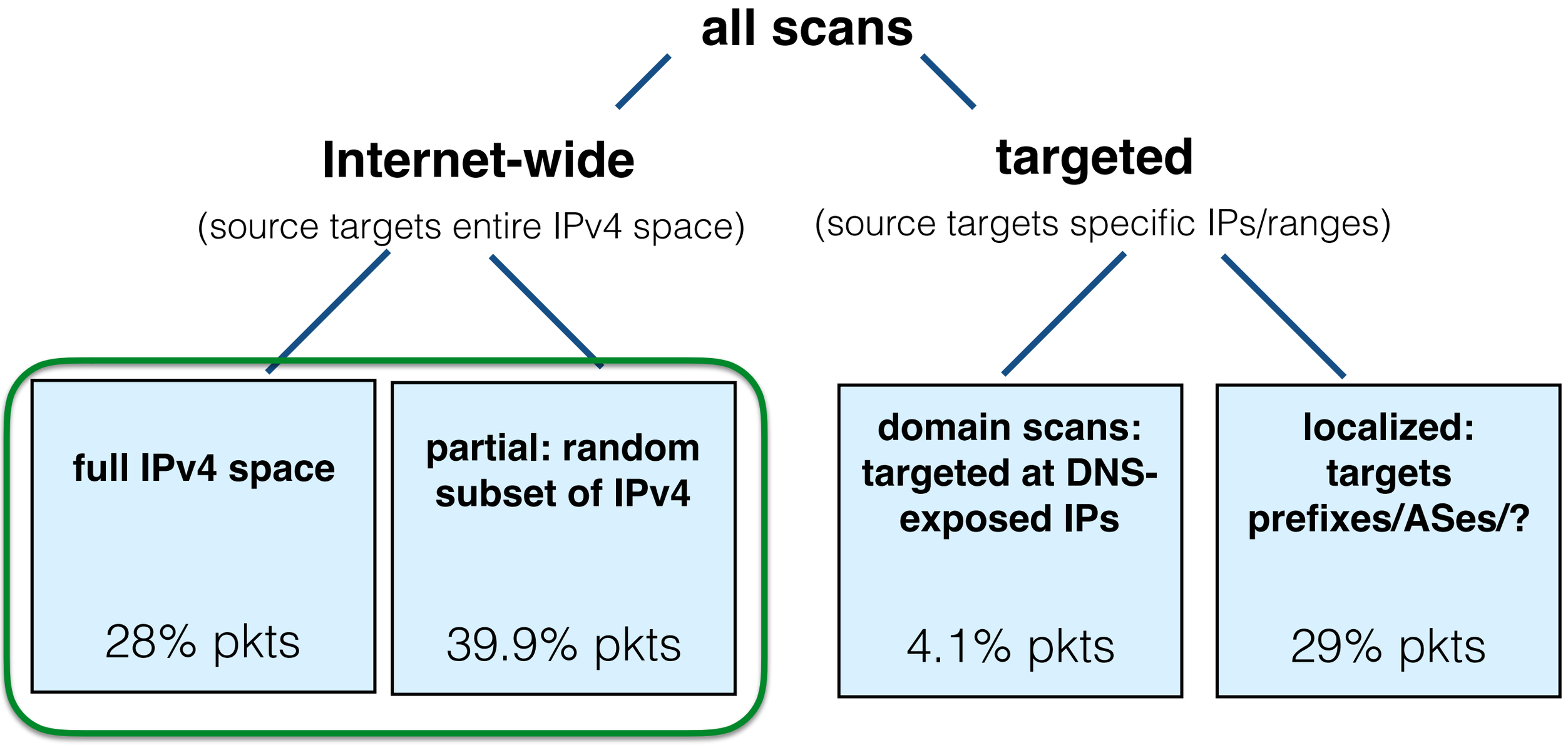
localized scans:
low pearson correlation

Internet-wide scans:
high pearson correlation

Scan Target Selection Strategies



Scan Target Selection Strategies



~68% scan packets: Internet-wide scans

Scan Target Selection Strategies

all scans

Internet-wide

(source targets entire IPv4 space)

targeted

(source targets specific IPs/ranges)

full IPv4 space

28% pkts

partial: random subset of IPv4

39.9% pkts

domain scans: targeted at DNS-exposed IPs

4.1% pkts

localized: targets prefixes/ASes/?

29% pkts

~68% scan packets: Internet-wide scans

~29% scan packets: target prefixes/ASes/?

Some 30% of all scan traffic localized!

do **not** target a random subset of the space

do **not** target the CDN's customers

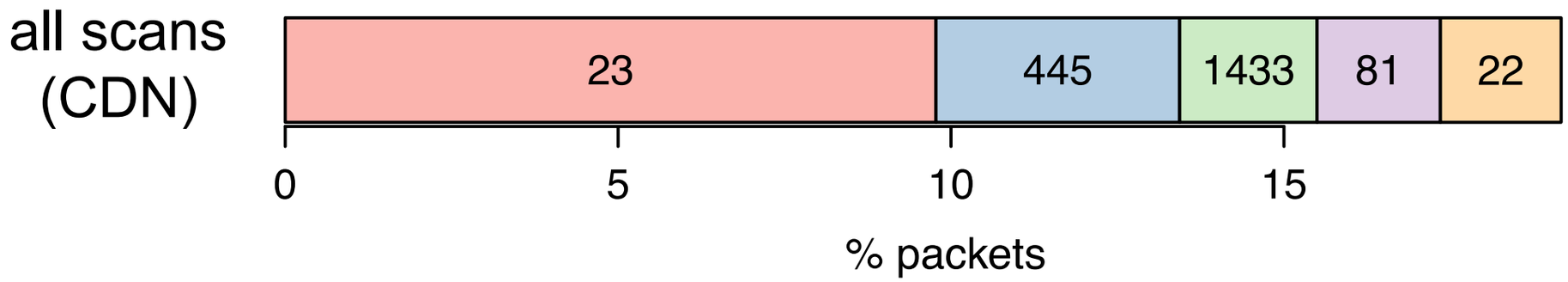
often target narrow regions of the IPv4 space (few prefixes)

invisible in darknets

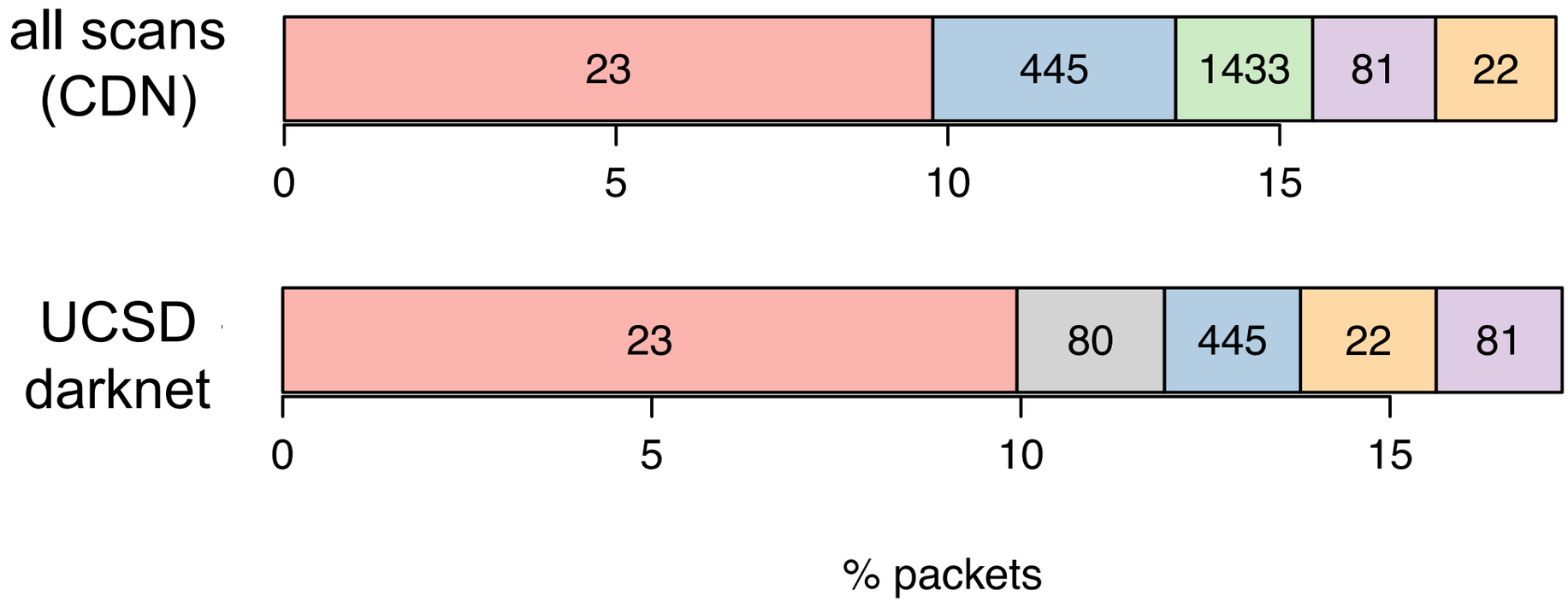
Vastly different characteristics:

Services scanned, Scanner origins, repeated scans, etc.

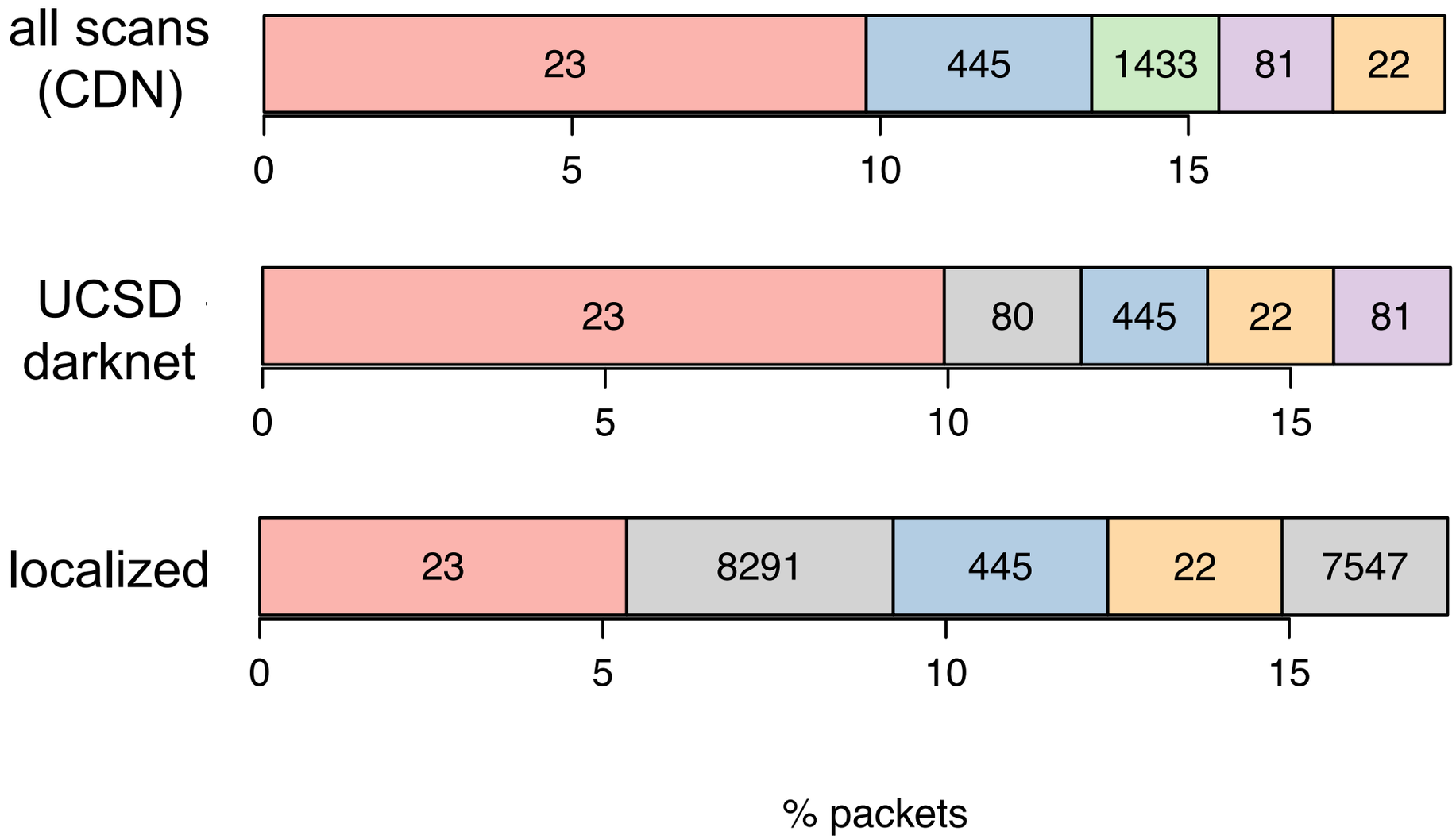
Top 5 Target Port Numbers



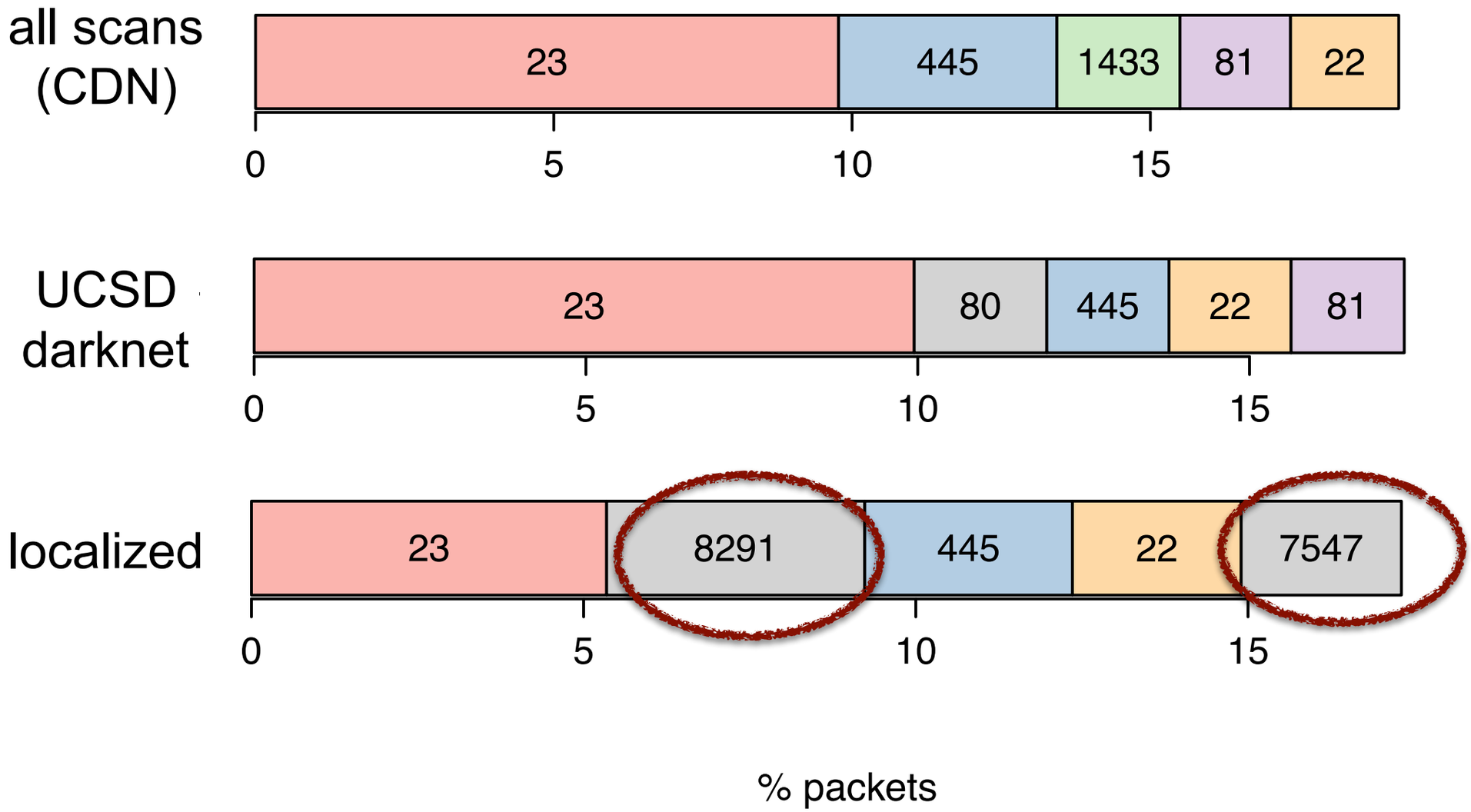
Top 5 Target Port Numbers



Top 5 Target Port Numbers



Top 5 Target Port Numbers



Critical vulnerabilities targeted in localized scans!
But barely scanned by Internet-wide scanners
-> Darknets severly underestimate scanning activity

Outline

Part 1

- What is a distributed network telescope?
- How much unsolicited traffic is there?

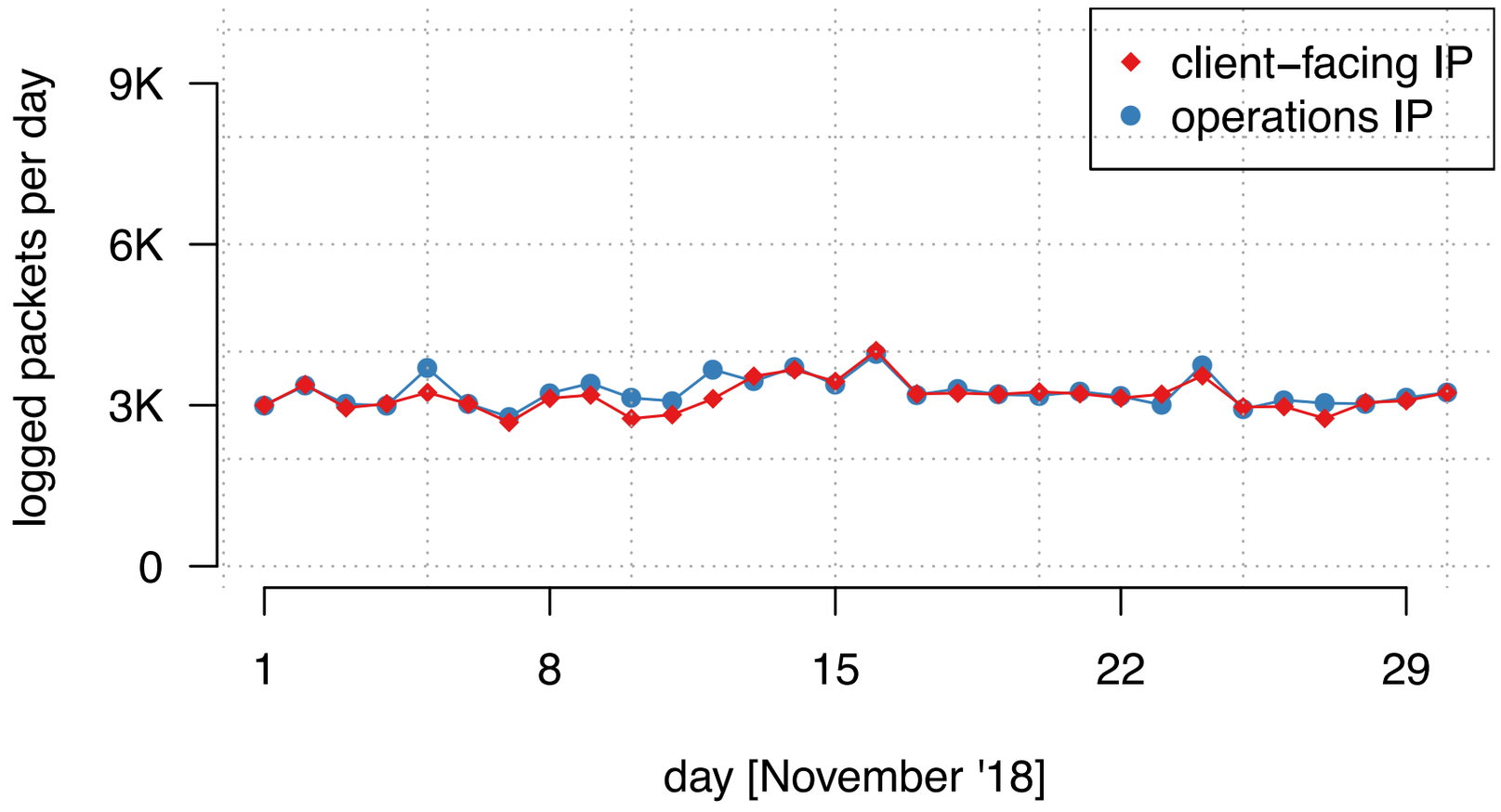
Part 2

- How much scanning traffic is there?
- How can we dissect scanning strategies?

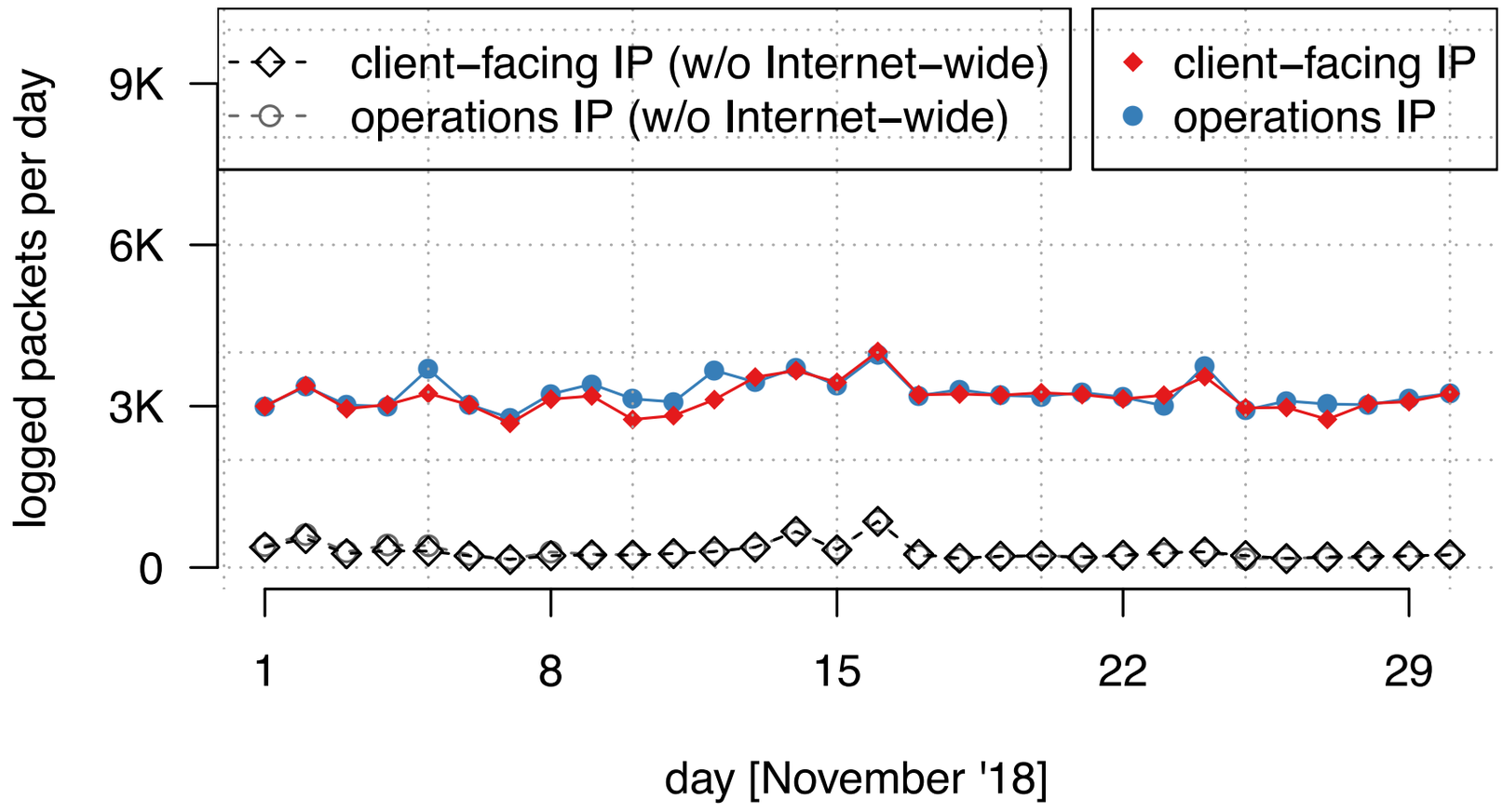
Part 3

- How can an individual network detect targeted scans?

Baseline Composition

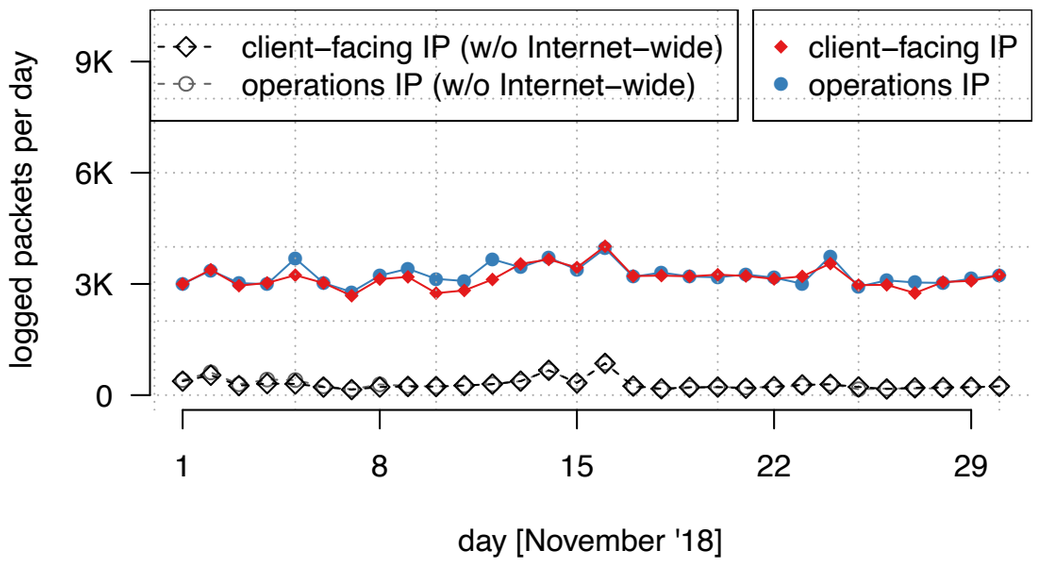


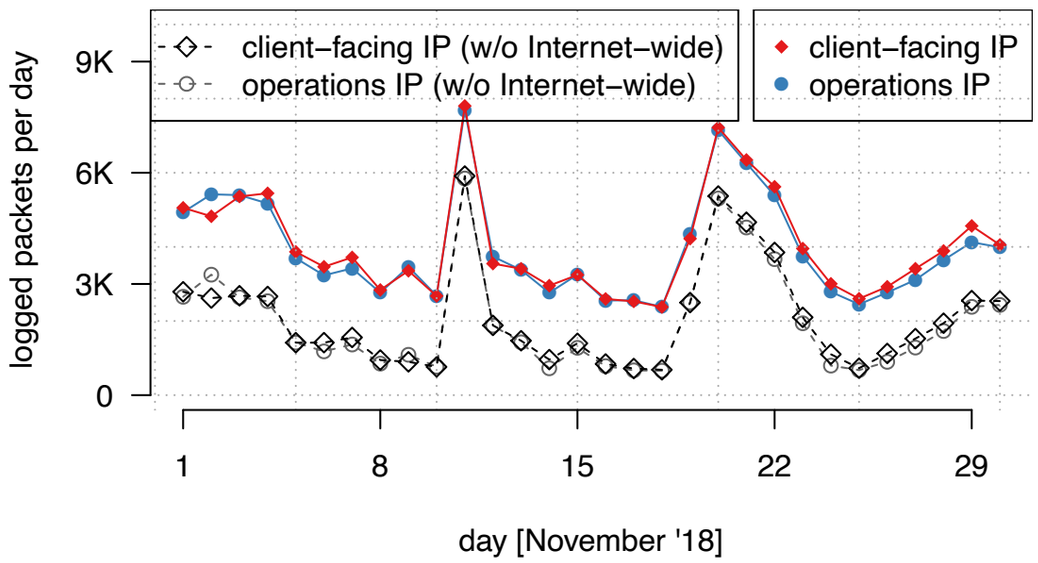
Baseline Composition



remaining unsolicited traffic
after removal of Internet-wide (full and random subset) scans

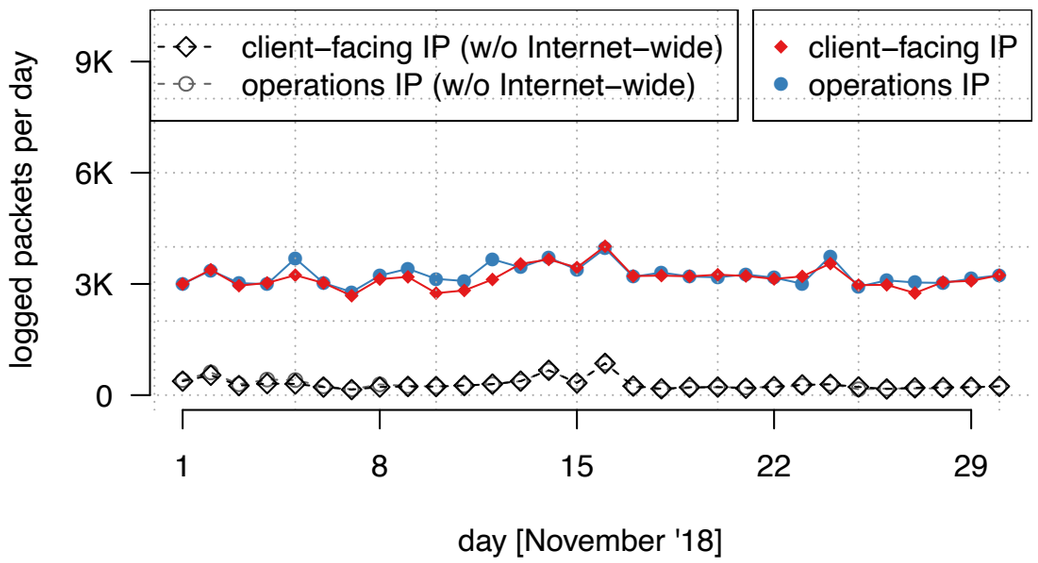
Baseline Composition

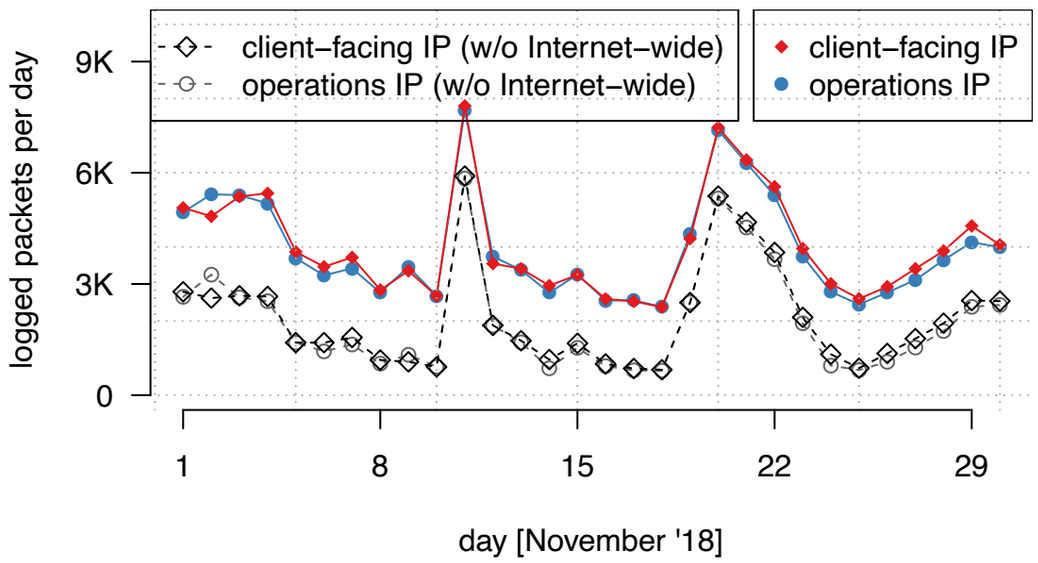




remaining unsolicited traffic
after removal of Internet-wide (full and random subset) scans

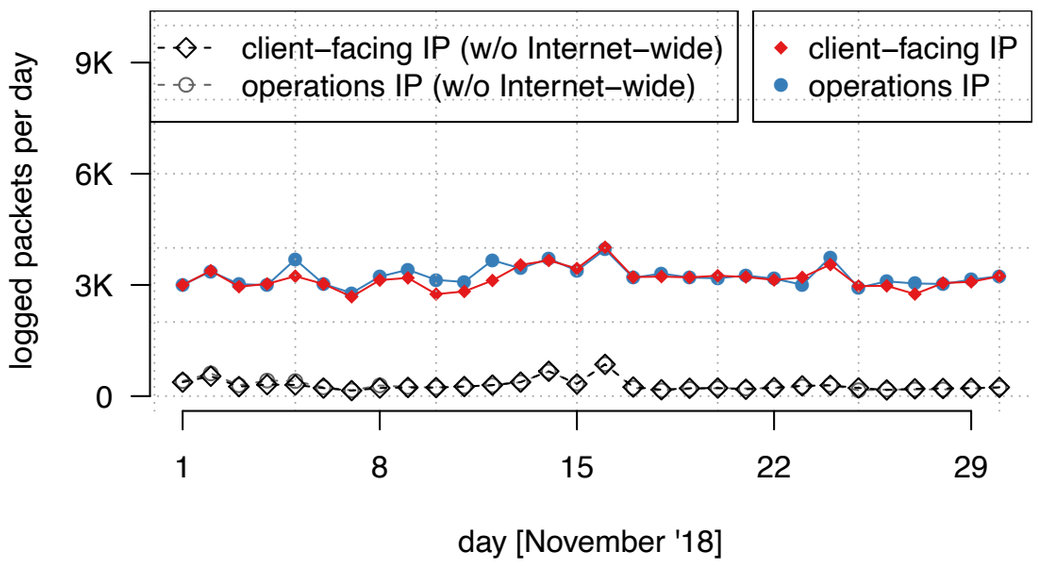
Baseline Composition

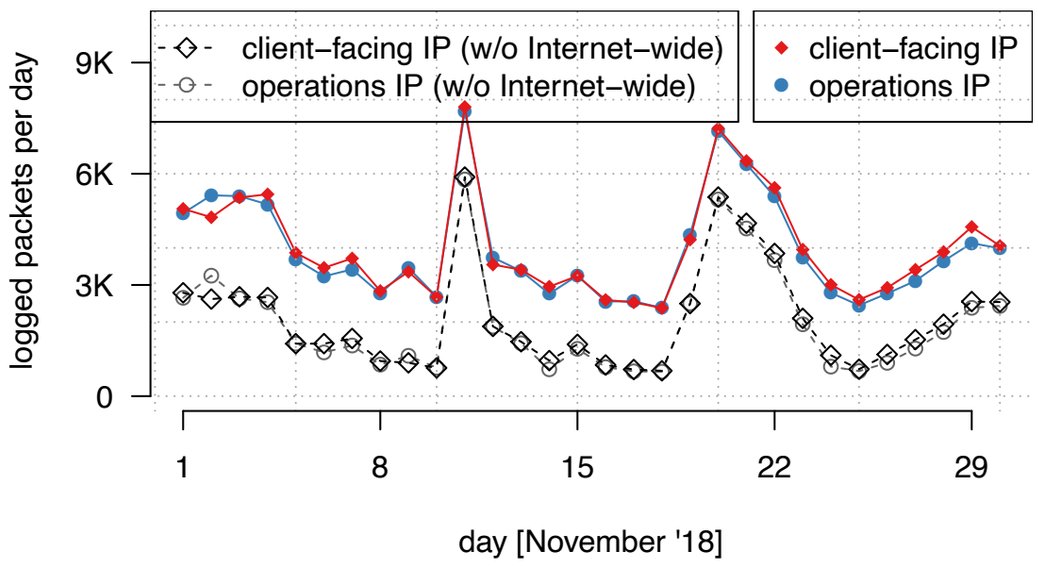




- 90% of baseline radiation is the result of Internet-wide scans
  - Full scans of the + random subsets of the IPv4 space

Baseline: Detect Targeted Scans



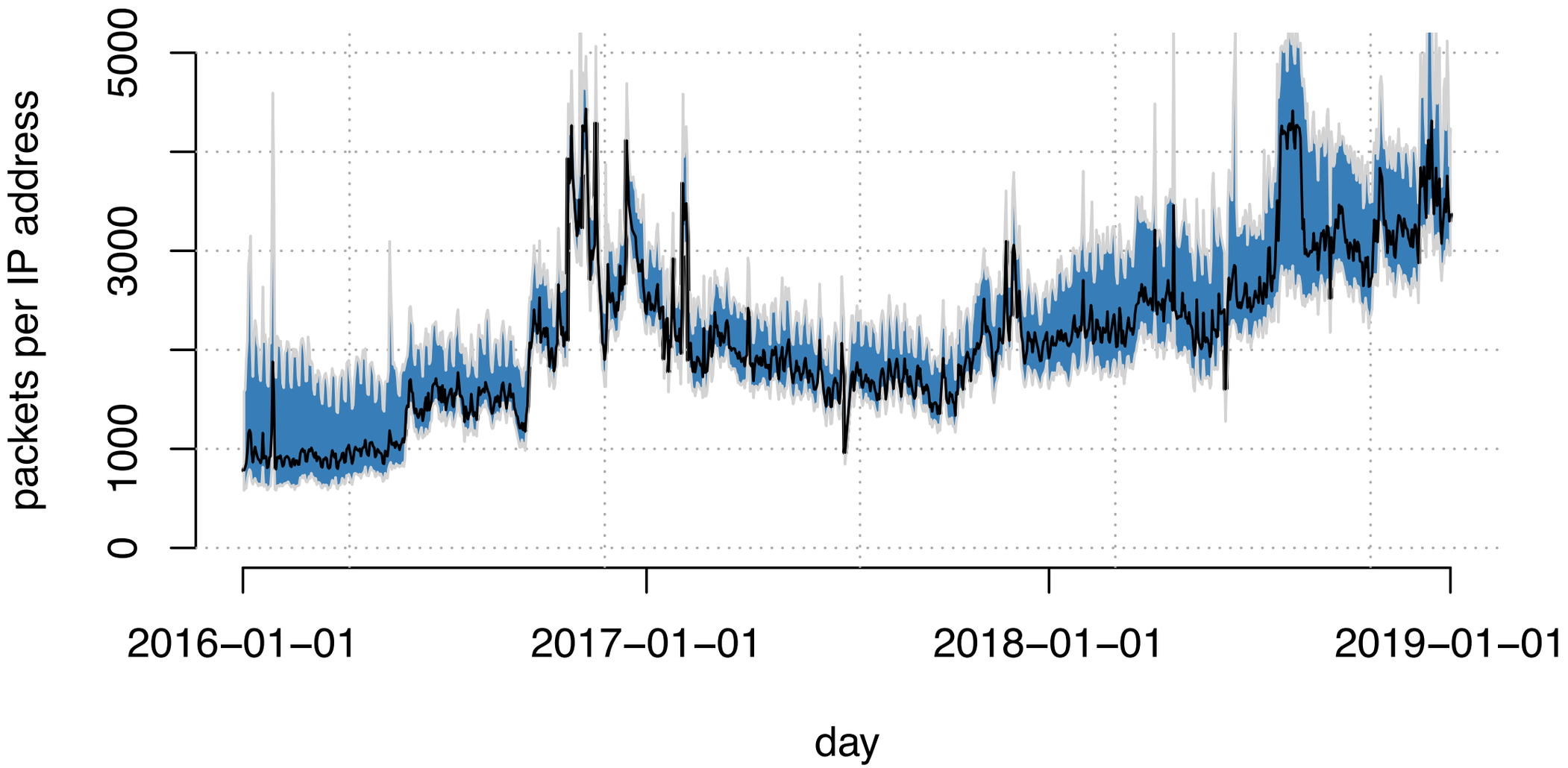
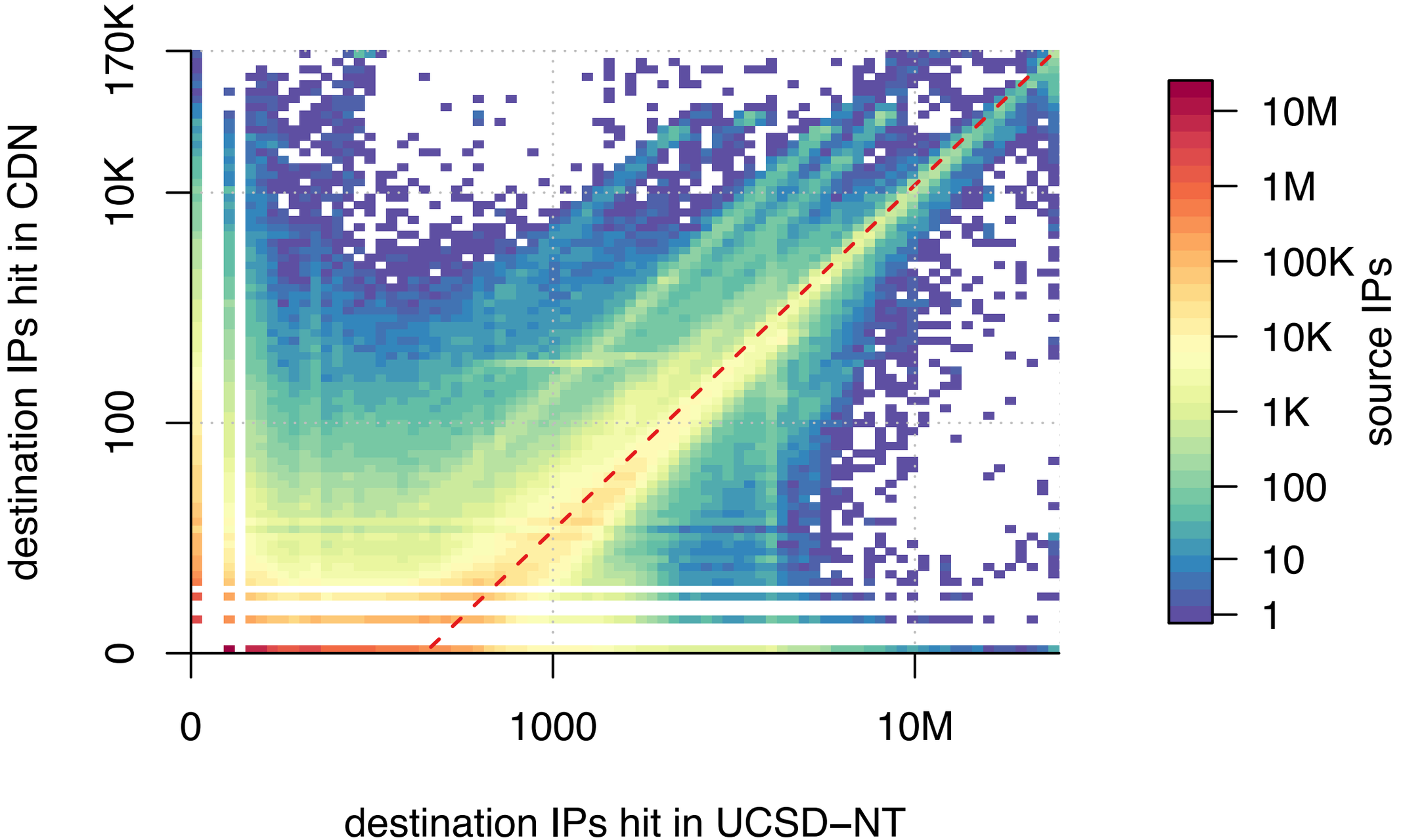


- 90% of baseline radiation is the result of Internet-wide scans
  - Full scans of the + random subsets of the IPv4 space
- Threat detection / detection of targeted activity
  - Determine current baseline level (darknets, etc.)
  - See if unsolicited scan traffic exceeds baseline

Key Takeaways

- Distributed telescope illuminates undocumented scan activity
- Baseline: Every routed IPv4 address receives ~3000pkts/day
  - Deviation from the baseline indicates targeted activity
- Methodology to dissect scan strategies
  - 30% of scan traffic are result of *localized* scans
  - Different characteristics (services targeted, origins, etc.)

Baseline Evolution



CDN vs. UCSD Darknet: DSTs per Source IP