Glowing in the Dark

Uncovering IPv6 Address Discovery and Scanning Strategies in the Wild

Hammas Bin Tanveer

In collaboration with:
Rachee Singh, Paul Pearce, Rishab Nithyanand

Scanning

sending unsolicited communication to an IP address **in order to draw a response**

- Measure protocol adoption
- Discover unadvertised services
- Analyze distributed systems
- Vulnerability scanning

Scanning in IPv6

- **IPv4 scanning**
  - Brute force scanning; scanning each and every one of the 2^32 addresses
  - Internet-wide scans in ~5 minutes with ZMap.

- IPv6 Scanning:
  - Brute force scanning is practically impossible now due to 2^128 addresses
  - Newer scanning techniques – this is what we characterize

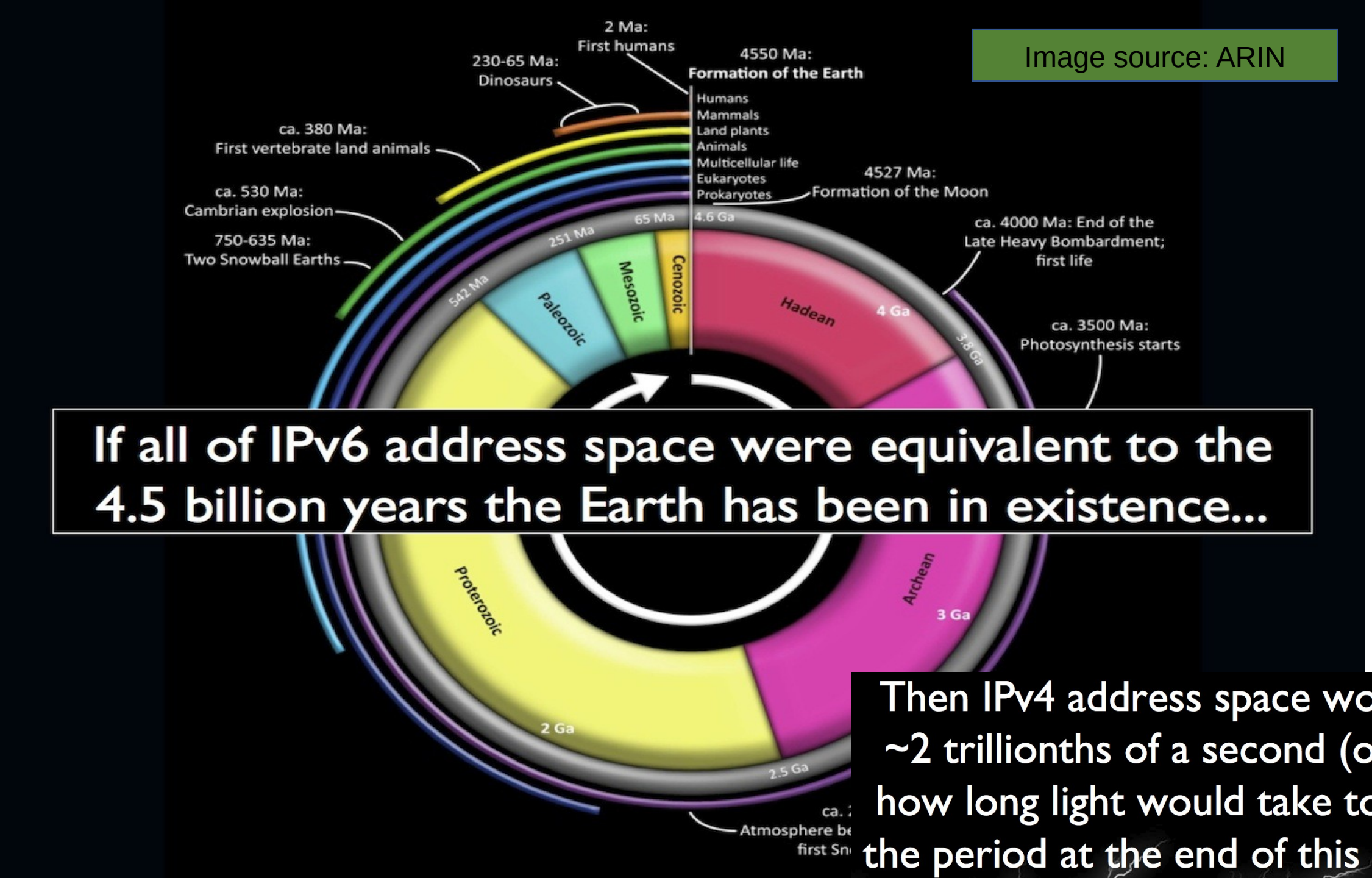


Image source: ARIN

If all of IPv6 address space were equivalent to the 4.5 billion years the Earth has been in existence...

Then IPv4 address space would equal ~2 trillionths of a second (or around how long light would take to traverse the period at the end of this sentence).

How to scan IPv6?

- IP Scanning
  - Pattern based
    - Collecting allocated IPv6 addresses
    - Learning patterns from these allocated addresses and generating “target” addresses to scan
    - Lower-byte addresses (RFC 7707)
- NXDOMAIN Scanning

How to scan IPv6?

- IP Scanning
  - Pattern based
    - Collecting allocated IPv6 addresses
    - Learning patterns from these allocated addresses and generating “target” addresses to scan
    - Exploiting IID addresses (RFC 7707)

Reducing the search space of unknown addresses by analyzing patterns in IP addresses

- NXDOMAIN Scanning

How to scan IPv6?

- IP Scanning
  - Pattern based
    - Collecting allocated IPv6 addresses
    - Learning patterns from these allocated addresses and generating "target" addresses to scan
    - Entropy/IP, addresses (RFC 7707)

Reducing the search space of unknown addresses by analyzing patterns in IP addresses

- NXDOMAIN Scanning

Reducing the search space of unknown addresses by exploiting semantics described in RFC 8020

How to scan IPv6?

- IP Scanning
  - Pattern based
    - Collecting allocated IPv6 addresses
    - Learning patterns from these allocated addresses and generating “target” addresses to scan
    - Lower-byte addresses (RFC 7707)

Reducing the search space of unknown addresses by analyzing patterns in IP addresses

Reduces the search space but still probabilistic

- NXDOMAIN Scanning

Reducing the search space of unknown addresses by exploiting semantics described in RFC 8020

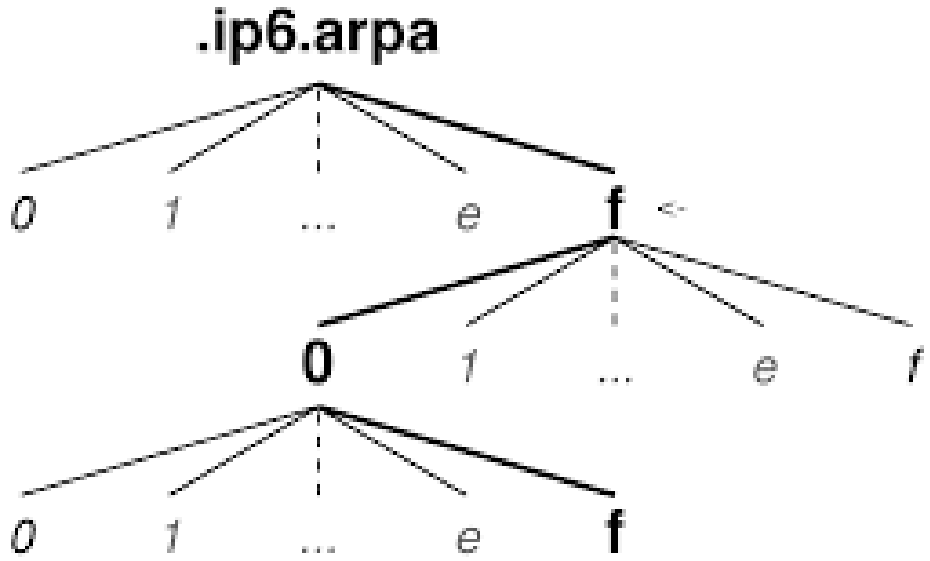
Will always return the correct allocated address

NXDOMAIN Scanning

RFC 8020 - **`NXDOMAIN: There Really Is Nothing Underneath`**

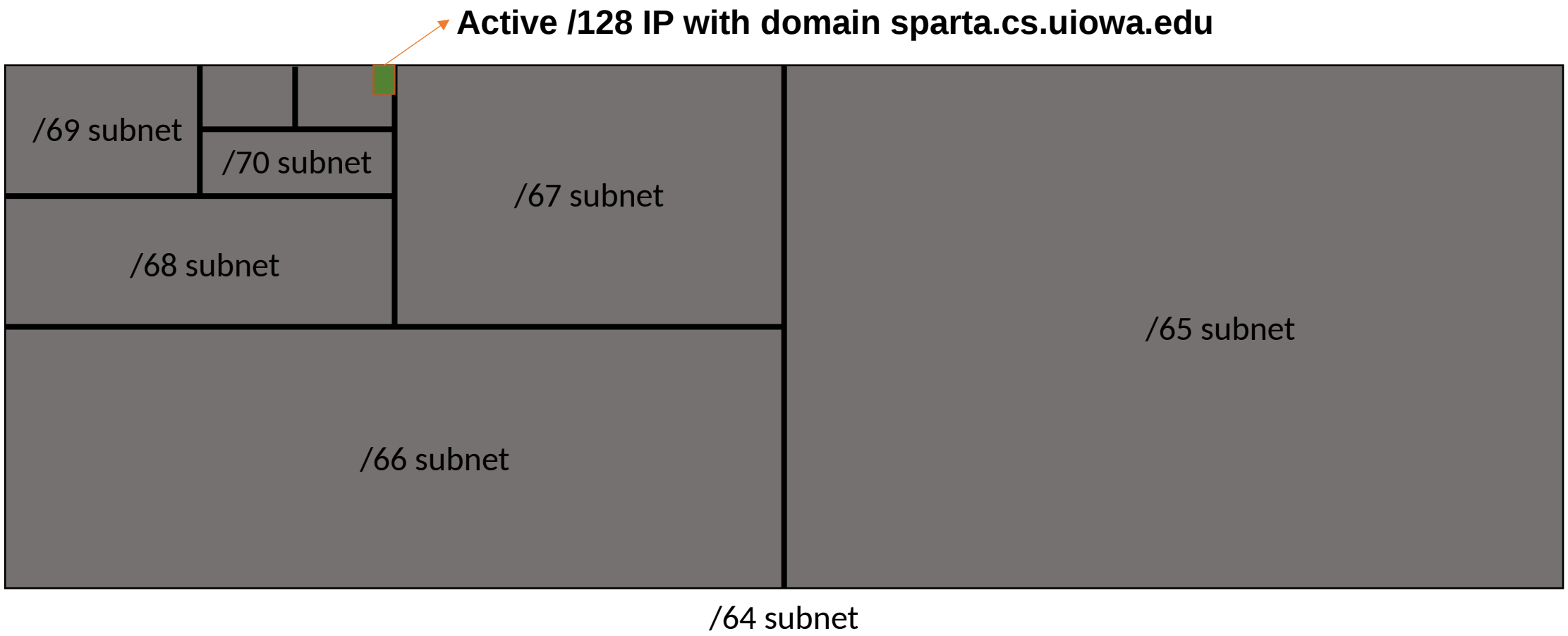
an NXDOMAIN response for a domain name means that no child domains underneath the queried name exist either

When this RFC is applied to DNS reverse trees, it unintentionally presents a side channel for efficient scanning of the IPv6 address space



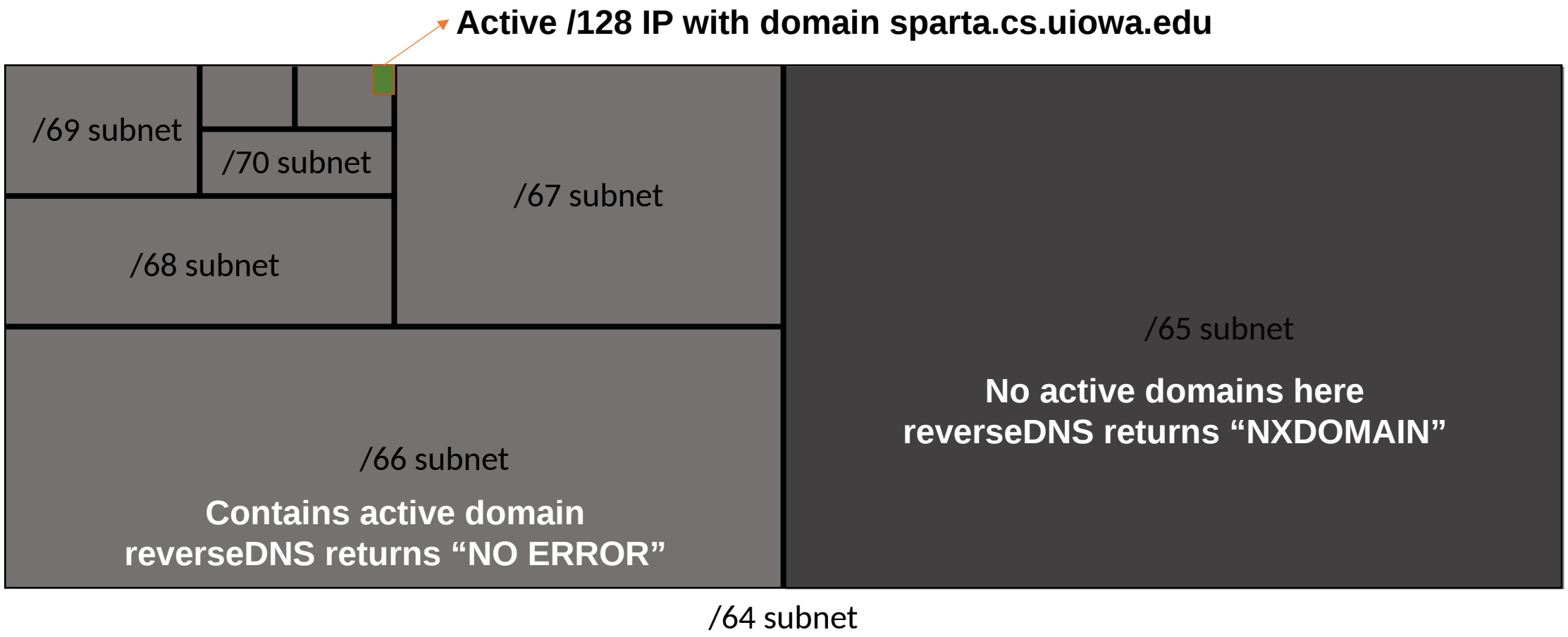
NXDOMAIN Scanning

RFC 8020 unintentionally presents a side channel for efficient scanning.



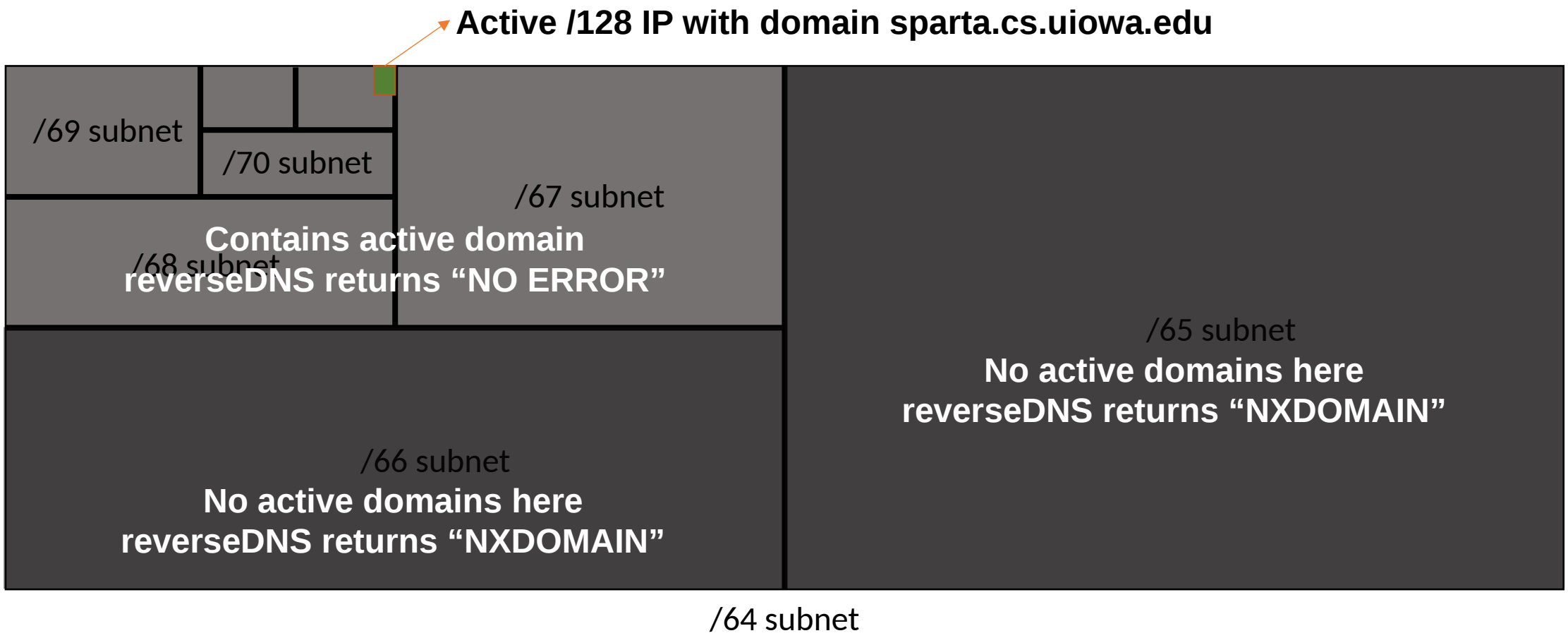
NXDOMAIN Scanning

RFC 8020 unintentionally presents a side channel for efficient scanning.



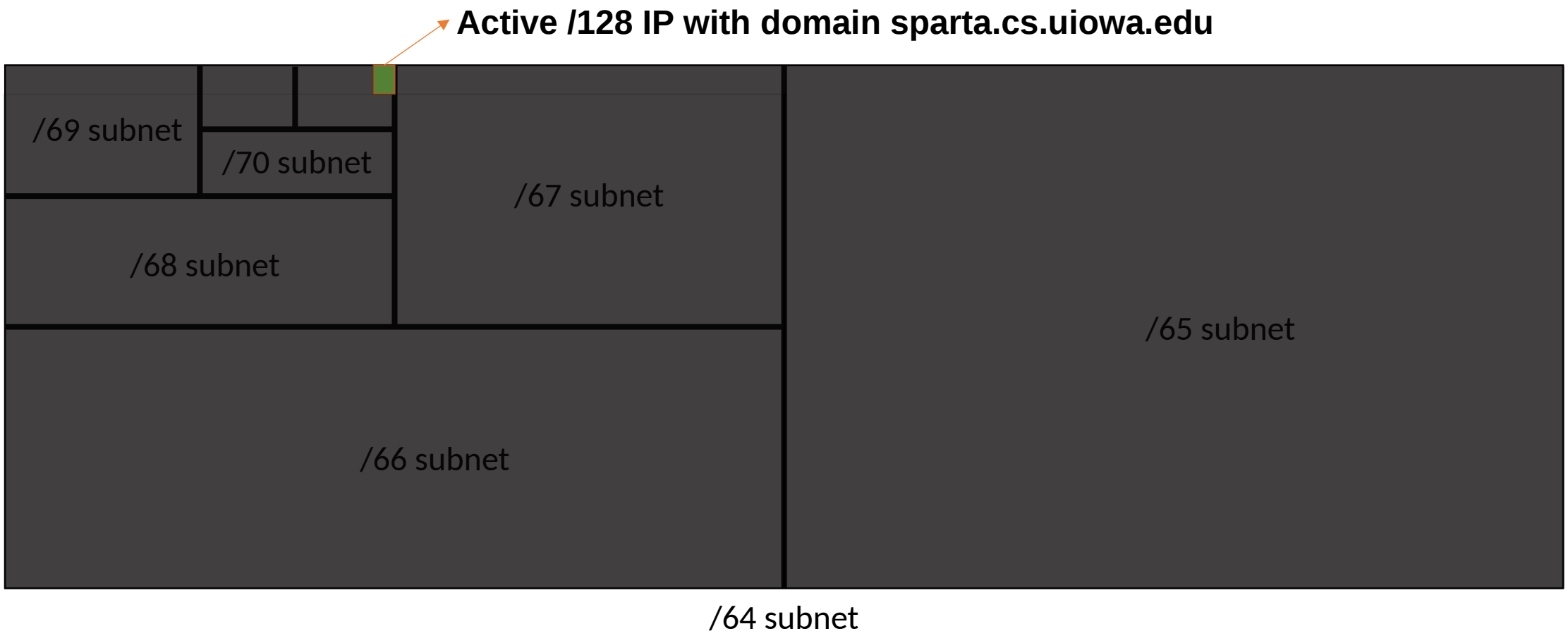
NXDOMAIN Scanning

RFC 8020 unintentionally presents a side channel for efficient scanning.



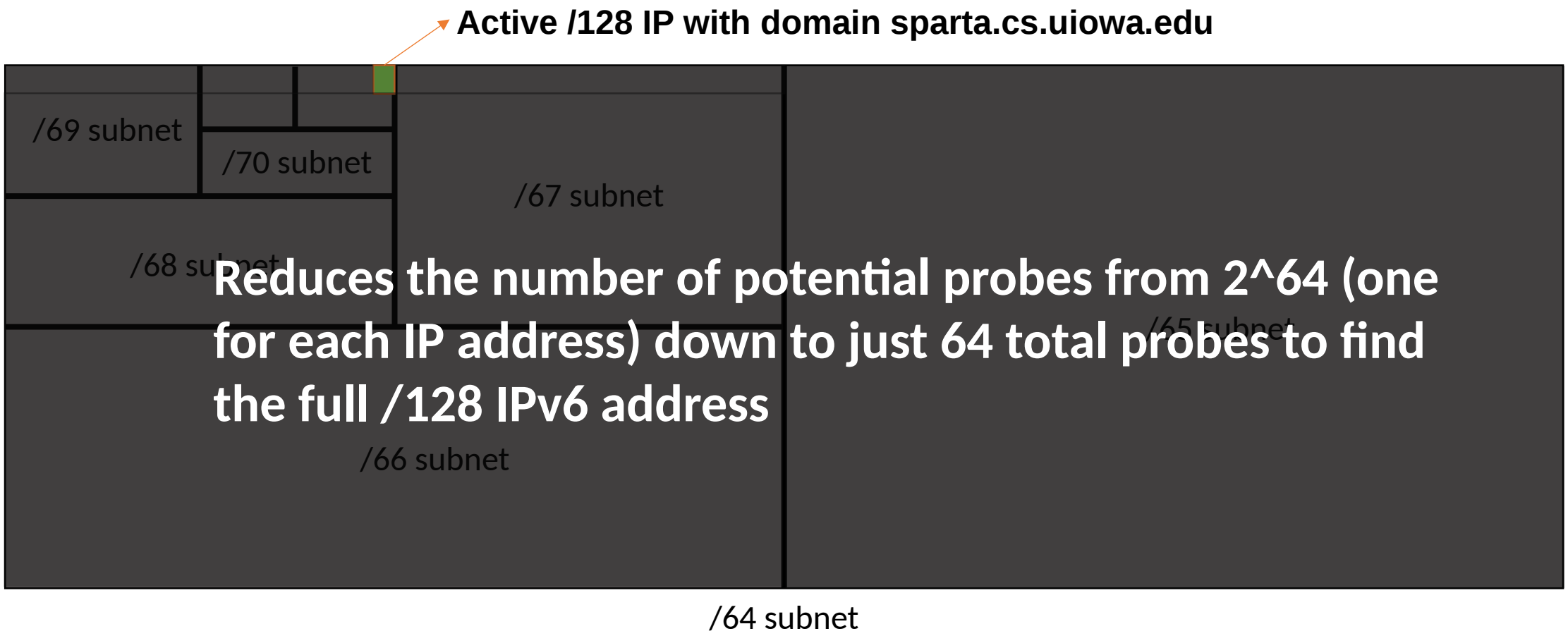
NXDOMAIN Scanning

RFC 8020 unintentionally presents a side channel for efficient scanning.



NXDOMAIN Scanning

RFC 8020 unintentionally presents a side channel for efficient scanning.



Reduces the number of potential probes from 2^64 (one for each IP address) down to just 64 total probes to find the full /128 IPv6 address

Experimental Setup

Experimental Setup

Goals:

- Mimic an active IPv6 address space
- Capture actual scanning traffic; both DNS scans and IP scans
- Link scanning activity to type of addresses and services deployed

Previous studies:

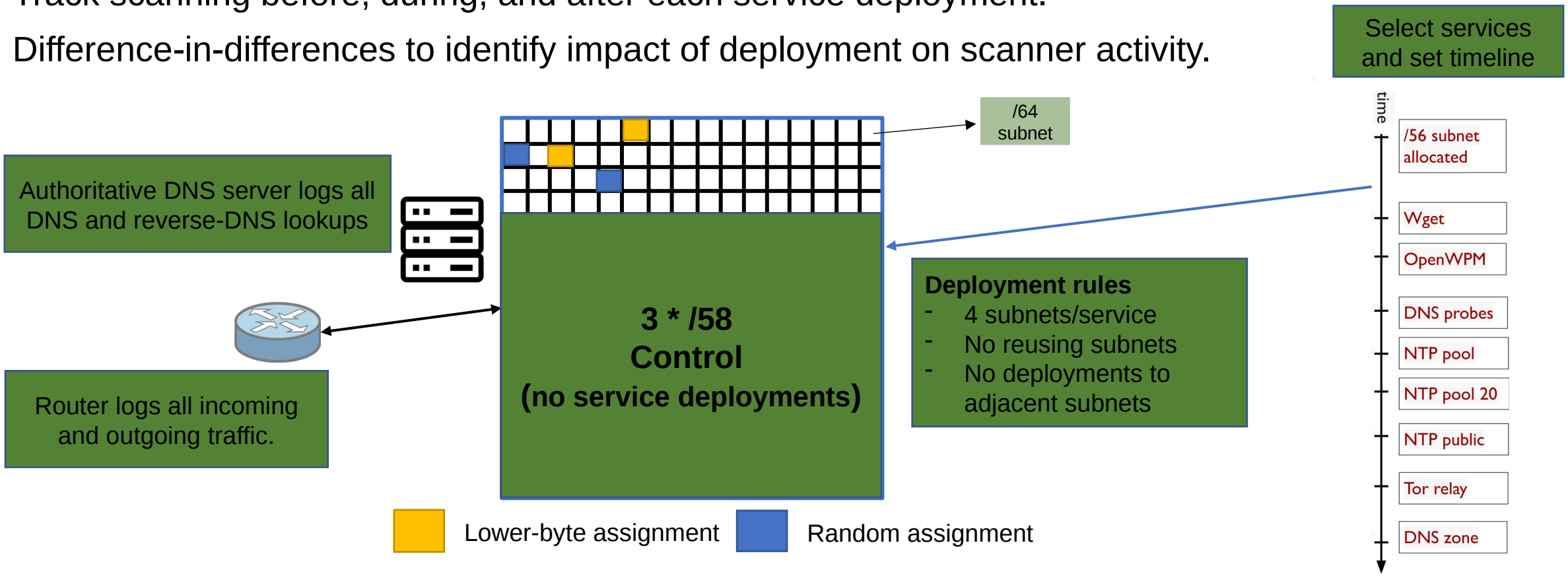
- Captured scanning traffic in a darknet; most of which was a result of IPv6 misconfigurations
- Captured scanning traffic for an authoritative DNS server; cannot link address types/services to scanning activity

Experimental setup

Start with a previously **unannounced** and **unused** /56 subnet (256 /64 subnets).

Track scanning before, during, and after each service deployment.

Difference-in-differences to identify impact of deployment on scanner activity.



observations

IPv6 scanning is common

		DNS logs	PCAPs
# Scanner probes	Before	200,335	13,044
	After	499,638	14,564,017
	Total	699,973	14,577,061
# Scanners	Before	20	5
	After	89	1065
	Total	96	1068

- We see some scanning activity even before any of our services were deployed

Key observations

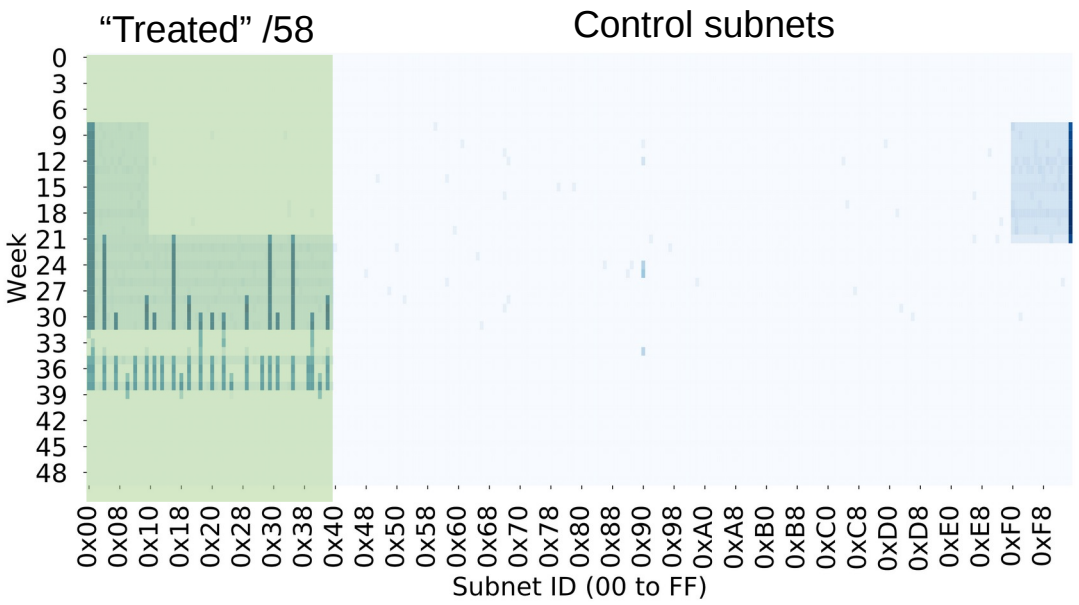
IPv6 scanning is common

		DNS logs	PCAPs
# Scanner probes	Before	200,335	13,044
	After	499,638	14,564,017
	Total	699,973	14,577,061
# Scanners	Before	20	5
	After	89	1065
	Total	96	1068

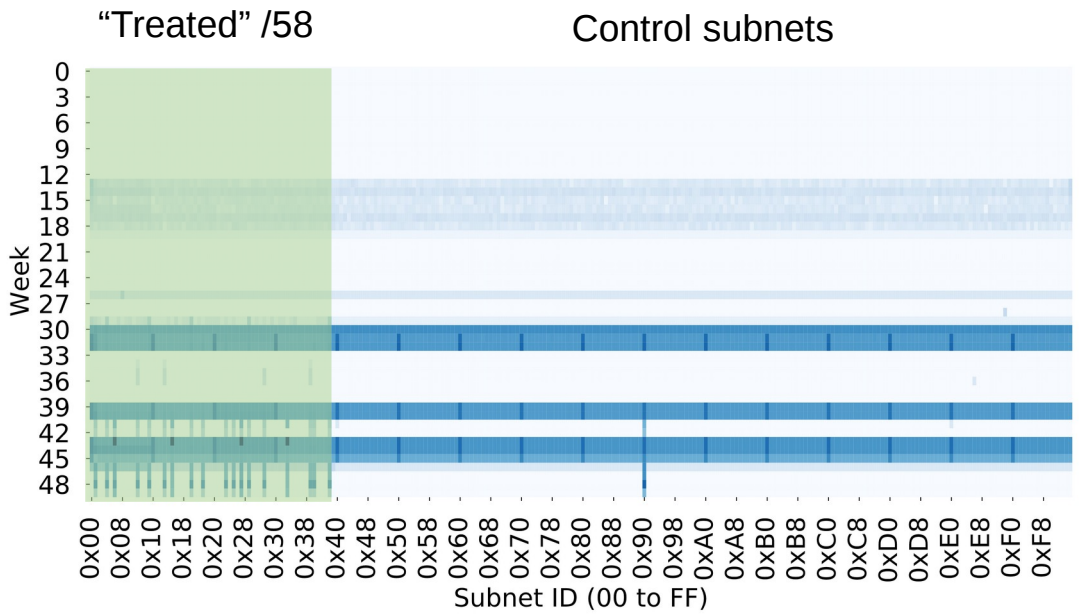
- We see some scanning activity even before any of our services were deployed
- Scanning activity increases significantly after services were deployed
  - in terms of both number of scanners and number of probes

observations

NXDOMAIN scanners are using the side channel efficiently



NXDOMAIN Scanners

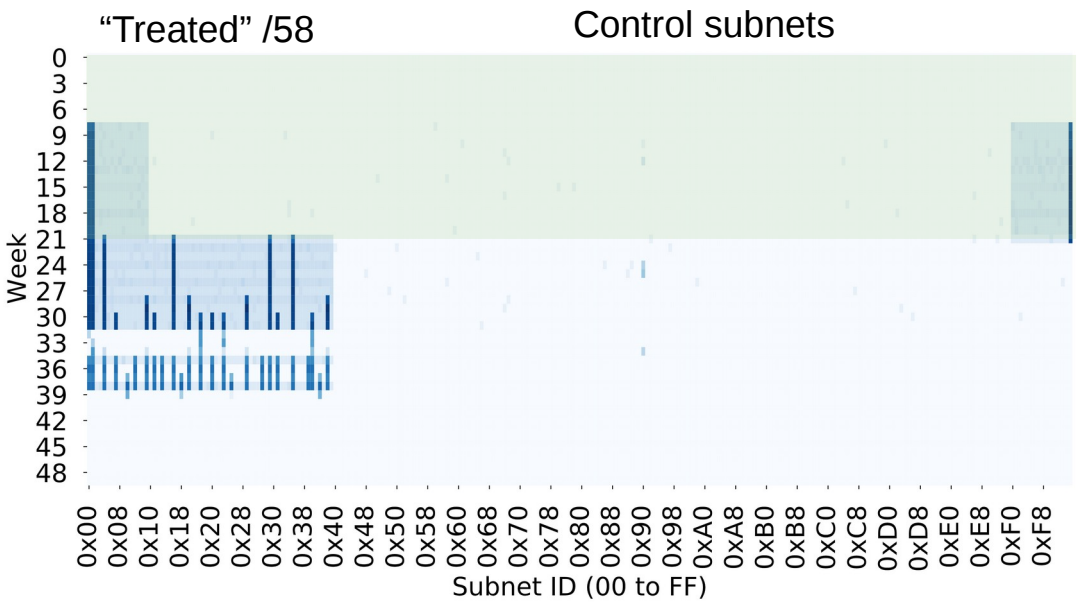


IP Scanners

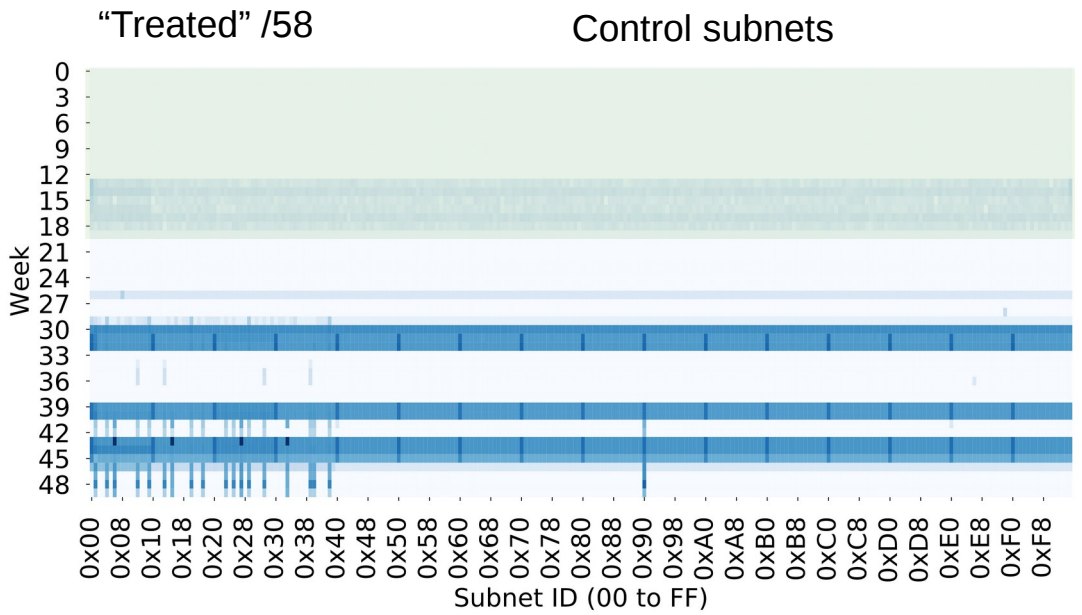
- The /58 used by our experimentation is highlighted
- Y-axis represents the entirety of our experimentation duration

Key observations

NXDOMAIN scanners are using the side channel efficiently



NXDOMAIN Scanners



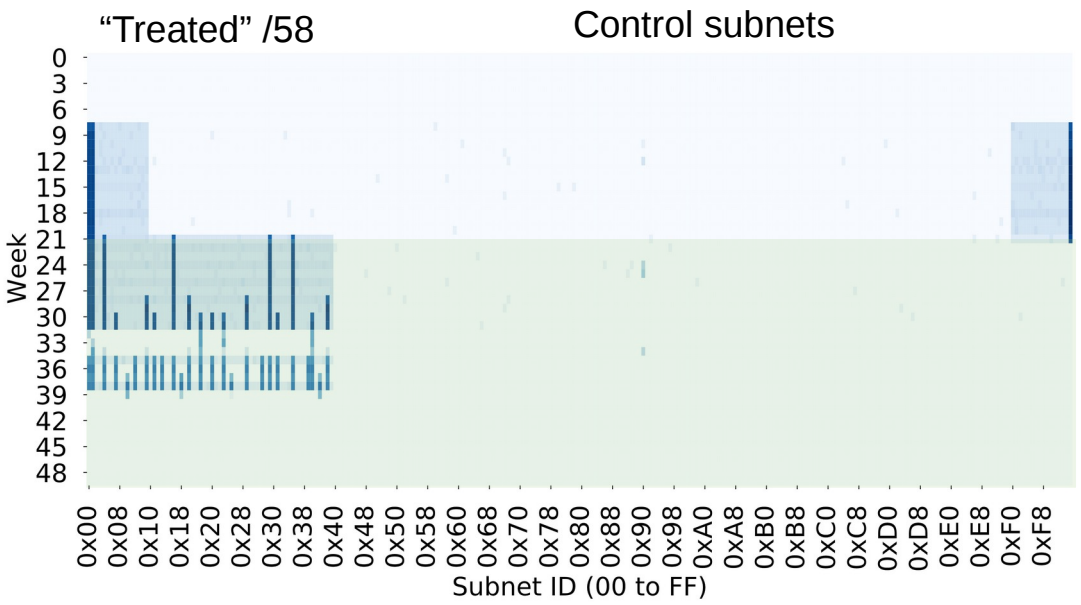
IP Scanners

The duration before any of our services started is highlighted

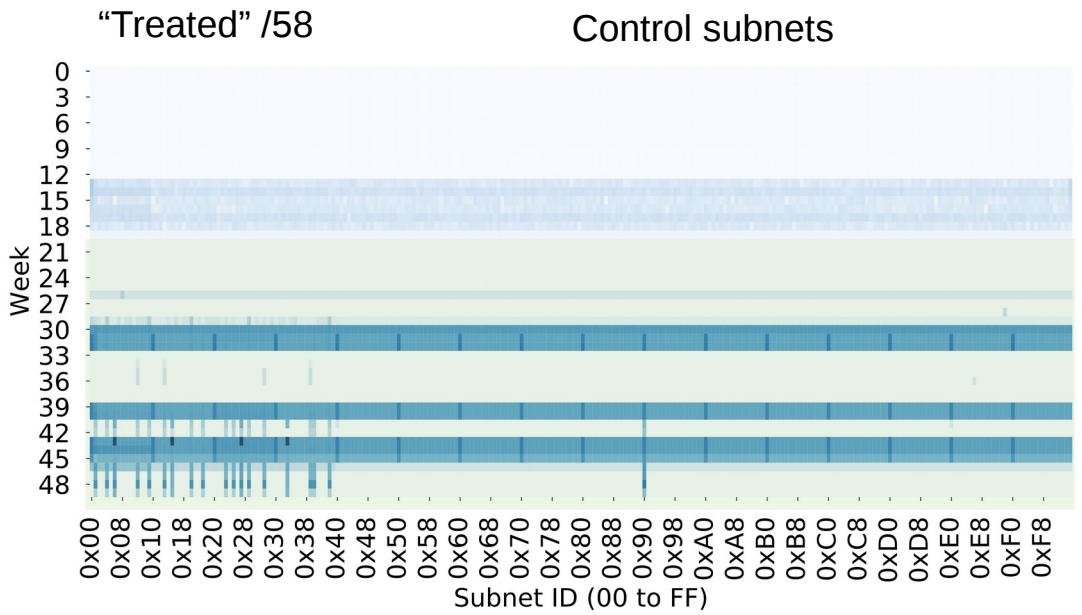
- NXDOMAIN scanners only scan the start and the end of our address space
- IP scanners scan the entirety of our subnet

Key observations

NXDOMAIN scanners are using the side channel efficiently



NXDOMAIN Scanners



IP Scanners

The duration after our services started is highlighted

- NXDOMAIN scanners stay strictly inside the /58 address space that hosted services during our experimentation; they are using the RFC 8020 side channel effectively
- IP scanners STILL scan the entirety of our subnet as they cannot leverage the side channel

Key observations

Service	DNS logs Δ_s^{diff}	DNS logs Δ_s^{C}	PCAPs Δ_s^{diff}	PCAPs Δ_s^{C}
wget	**511.1**	-2.9	0.0	0.0
OpenWPM	**564.0**	-0.1	-0.4	0.4
DNS probes	**736.0**	1.9	**128.8**	**265.1**
NTP_{pool}	**348.3**	**6.0**	**313.6**	**734.4**
$\text{NTP}_{pool-20}$	**6.5**	**-9.7**	636.9	0.0
NTP_{public}	**72.2**	1.6	**116.3**	0.0
Tor relay	**87.1**	-0.4	0.0	0.0
DNS Zone	**1.2**	-1.3	-54.4	**588.1**

- NXDOMAIN scanners target treatment subnets for all the services; they target control subnets much less

Δ_s^{diff} Increase in scanning activity within the treatment subnets where the services were running

$\Delta_{s,t}^{C}$ Increase in scanning activity in the control subnets i.e where no services were running

Bold numbers represent statistically significant differences

Key observations

Service	DNS logs Δ_s^{diff}	DNS logs Δ_s^{C}	PCAPs Δ_s^{diff}	PCAPs Δ_s^{C}
wget	**511.1**	-2.9	0.0	0.0
OpenWPM	**564.0**	-0.1	-0.4	0.4
DNS probes	**736.0**	1.9	**128.8**	**265.1**
NTP_{pool}	**348.3**	**6.0**	**313.6**	**734.4**
$NTP_{pool-20}$	**6.5**	**-9.7**	636.9	0.0
NTP_{public}	**72.2**	1.6	**116.3**	0.0
Tor relay	**87.1**	-0.4	0.0	0.0
DNS Zone	**1.2**	-1.3	-54.4	**588.1**

- NXDOMAIN scanners target treatment subnets for all the services; they target control subnets much less
- IP scanners exhibit mixed behavior; they target both, treatment and control subnets

Δ_s^{diff} Increase in scanning activity within the treatment subnets where the services were running

$\Delta_{s,t}^{C}$ Increase in scanning activity in the control subnets i.e where no services were running

Bold numbers represent statistically significant differences

Key observations

Service	DNS logs Δ_s^{diff}	DNS logs Δ_s^{C}	PCAPs Δ_s^{diff}	PCAPs Δ_s^{C}
wget	**511.1**	-2.9	0.0	0.0
OpenWPM	**564.0**	-0.1	-0.4	0.4
DNS probes	**736.0**	1.9	**128.8**	**265.1**
NTP_{pool}	**348.3**	**6.0**	**313.6**	**734.4**
$NTP_{pool-20}$	**6.5**	**-9.7**	636.9	0.0
NTP_{public}	**72.2**	1.6	**116.3**	0.0
Tor relay	**87.1**	-0.4	0.0	0.0
DNS Zone	**1.2**	-1.3	-54.4	**588.1**

- NXDOMAIN scanners target treatment subnets for all the services; they target control subnets much less
- IP scanners exhibit mixed behavior; they target both, treatment and control subnets
- NXDOMAIN scanners target different services than IP scanners

Δ_s^{diff} Increase in scanning activity within the treatment subnets where the services were running

$\Delta_{s,t}^{C}$ Increase in scanning activity in the control subnets i.e where no services were running

Bold numbers represent statistically significant differences

Key takeaways

Analyzing dark traffic is not the best way to study scanner behavior.

- Particularly in IPv6 where scanners need to be intelligent to succeed.

Most scanners aren’t using the state-of-the-art (NXDOMAIN scanning), but its only a matter of time.

- Is the efficiency from NXDOMAIN responses worth the loss of defense against scanning? [RFC8020]

Address discovery methods are very different than we expected.

- Public lists of IPs (e.g., Tor, NTP, Zone) were used significantly less than data from DNS resolvers.

Neighboring networks should expect scanning activity.

- Most IP scanners are broad scanners which, on average, spend more time probing immediate neighbors of active /64s.

Questions and comments