Packed to the Brim: Investigating the Impact of Highly Responsive Prefixes on Internet-wide Measurement Campaigns

PATRICK SATTLER, Technical University of Munich, Germany
JOHANNES ZIRNGIBL, Technical University of Munich, Germany
MATTIJS JONKER, University of Twente, The Netherlands
OLIVER GASSER, Max Planck Institute for Informatics, Germany
GEORG CARLE, Technical University of Munich, Germany
RALPH HOLZ, University of Münster, Germany

Internet-wide scans are an important tool to evaluate the deployment of services. To enable large-scale application layer scans, a fast, stateless port scan (*e.g.*, using ZMap) is often performed ahead of time to collect responsive targets. It is a common expectation that port scans on the entire IPv4 address space provide a relatively unbiased view as they cover the complete address space. Previous work, however, has found prefixes where all addresses share particular properties. In IPv6, aliased prefixes and fully responsive prefixes, *i.e.*, prefixes where all addresses are responsive, are a well-known phenomenon. However, there is no such in-depth analysis for prefixes with these responsiveness patterns in IPv4.

This paper delves into the underlying factors of this phenomenon in the context of IPv4 and evaluates port scans on a total of 161 ports (142 TCP & 19 UDP ports) from three different vantage points. To account for packet loss and other scanning artifacts, we propose the notion of a new category of prefixes, which we call highly responsive prefixes (HRPs). Our findings show that the share of HRPs can make up 70 % of responsive addresses on selected ports. Regarding specific ports, we observe that CDNs contribute to the largest fraction of HRPs on TCP/80 and TCP/443, while TCP proxies emerge as the primary cause of HRPs on other ports. Our analysis also reveals that application layer handshakes to targets outside HRPs are, depending on the chosen service, up to three times more likely to be successful compared to handshakes with targets located in HRPs. To improve future scanning campaigns conducted by the research community, we make our study's data publicly available and provide a tool for detecting HRPs. Furthermore, we propose an approach for a more efficient, ethical, and sustainable application layer target selection. We demonstrate that our approach has the potential to reduce the number of TLS handshakes by up to 75 % during an Internet-wide scan while successfully obtaining 99 % of all unique certificates.

CCS Concepts: • **Networks** → *Middle boxes / network appliances*; *Network measurement*; *Public Internet*.

Additional Key Words and Phrases: highly responsive prefix, port scanning, ethical scanning

Authors' addresses: Patrick Sattler, sattler@net.in.tum.de, Technical University of Munich, Munich, Germany; Johannes Zirngibl, zirngibl@net.in.tum.de, Technical University of Munich, Munich, Germany; Mattijs Jonker, m.jonker@utwente.nl, University of Twente, Enschede, The Netherlands; Oliver Gasser, oliver.gasser@mpi-inf.mpg.de, Max Planck Institute for Informatics, Saarbrücken, Germany; Georg Carle, carle@net.in.tum.de, Technical University of Munich, Munich, Germany; Ralph Holz, ralph.holz@uni-muenster.de, University of Münster, Münster, Germany.

Proc. ACM Netw., Vol. 1, No. CoNEXT3, Article 24. Publication date: December 2023.

24

1 INTRODUCTION

In the last decade, Internet-wide port scans have been frequently used by network and security researchers to quantify service deployment or as the first step in a scanning pipeline that targets higher-layer protocols, *e.g.*, by chaining ZMap [13] and ZGrab2 [32]. Researchers routinely assess the impact of a service vulnerability with the help of port scans [21, 36, 37]. An overlooked bias can easily distort any conclusions drawn; therefore, high data quality is critical.

Previous research, *e.g.*, by Izhikevich et al. [23, 24] and Gasser et al. [16], identified a phenomenon where all IP addresses within a prefix seem to respond on a scanned port. In the case of IPv6, this type of prefix is particularly important to consider. The size of allocated IPv6 prefixes is such that a full scan is infeasible, and hence hitlists are commonly used. If such prefixes appear on a hitlist, they are likely to introduce an undesired bias in the scan results [17]. Analysis of data from IPv6 scans has revealed several possible reasons for the presence of such prefixes. One of these reasons is the existence of aliased prefixes, which are prefixes where a single host responds to all addresses within that prefix [16]. Further work by Zirngibl et al. [43] in 2022 found that some Content Delivery Network (CDN) prefixes appear as fully responsive. Although such prefixes appear to be aliased, no single responder is behind an address, either. Therefore, the authors introduced the fully responsive prefix (FRP) class, which generalizes the concept of aliased prefixes in IPv6.

For IPv4, however, FRPs are commonly not further considered by researchers. In the case of SYN scans, scanners are fast enough to iterate over all IPv4 addresses. However, this does not apply to application-layer scans, which are typically resource-heavy and take more time. Here, FRPs can be much more problematic, especially if the deployment of the targeted protocol is low or if the same host responds for many addresses. An application-layer scan that includes FRPs will take much longer than needed and may introduce bias or lead to unwanted artifacts in the data. Consequently, security assessments or the evaluation of protocol deployments can lead to incorrect conclusions. Limiting port scans to the necessary volume is also ethically responsible, which is an additional incentive to avoid scanning FRPs.

By definition, FRPs exclusively describe prefixes where every single address within a prefix exhibits responsiveness on a specific protocol-port combination. In contrast to IPv6, where only a sample of addresses is probed to determine full responsiveness, in IPv4, we have the capability to test an entire prefix comprehensively. Since confounding factors—such as packet loss—may impact this probing of an entire prefix, we find the need for a new prefix classification that takes these factors into account. In this paper, we introduce highly responsive prefixes (HRPs), which are prefixes where over 90 % of addresses demonstrate responsiveness for a specific port. To the best of our knowledge, our study is the first of its kind concerning HRPs in IPv4.

Motivating Example: Our interest in this subject stemmed from curiosity about the impact of observed measurement artifacts (i.e., HRPs). When analyzing long-running measurement campaigns, we identified certain inefficiencies associated with the conventional approach of utilizing ZMap [13] for the detection of open ports, followed by an application layer scan on the successful addresses.

We conduct a weekly Transport Layer Security (TLS) scanning campaign to gather extensive TLS deployment information. Our approach is quite typical, commencing with a preliminary ZMap scan, followed by the application-layer scan using TLS. During this procedure, ZMap typically reports approximately 50 M IP addresses with TCP/443 port open. Of these, roughly 30 % are located within HRPs, with only 1.7 % of Autonomous Systems (ASes) responsible for all such HRPs. Therefore, repeatedly scanning the same ASes often fails to yield additional information, while potentially imposing a burden on infrastructure providers. Our initial analysis showed that CDNs play a significant, albeit not an exclusive role, which prompted us to investigate more in depth.

Durumeric et al. [13], proposed to use ZMap in the manner mentioned—namely, as a tool for identifying responsive hosts for application-layer scans and for determining service deployment statistics. However, our findings illustrate why this perspective may be overly simplistic and no longer entirely justified. In Section 2, we provide a list of some exemplary measurement studies that have been influenced by the existence of HRPs, and evaluate the available data.

Contributions: In this paper, we conduct an extensive analysis of HRPs in IPv4. To the best of our knowledge, there has been no prior study focused on HRPs in IPv4. We shed light on the presence and characteristics of HRPs through IPv4-based port scans from multiple vantage points. Additionally, we present an ethical scanning approach that takes into account the existence of HRPs. In this work, we make the following contributions:

(i) We conduct a comprehensive analysis of highly responsive prefixes in IPv4 across 161 ports. Our findings reveal that HRPs consistently cover approximately 30 % of responsive addresses on TCP ports 80 and 443. Furthermore, this coverage can reach up to 70 % for less common ports.

(ii) We evaluate HRPs from both application-layer and Domain Name System (DNS) perspectives. Our analysis shows that targets outside HRPs exhibit up to three times higher responsiveness at the application layer compared to targets inside HRPs.

(iii) We provide a tool for detecting HRPs based on port scan output, such as that generated by ZMap. Additionally, we publish detailed statistics for all ports analyzed [33] and offer continuous weekly HRP results for TCP ports 80 and 443 on our website – **https://hrp-stats.github.io/**.

(iv) We discuss our findings and the potential for more ethical scanning practices in the future. Our proposed adjusted scanning approach enhances scan success rates by selectively targeting specific hosts within HRPs, thereby reducing the load on scanned infrastructure. Our results demonstrate that, with this approach, we can decrease the number of application-layer scans by up to 70 % while retaining all essential information about the target prefixes.

Outline: Related work is covered in Section 2. In Section 3, we introduce the conducted scans and datasets we used. In Section 4, we define highly responsive prefixes and explain our approach and tooling to detect them. We evaluate the presence, stability, and specifics of HRPs in Section 5. Finally, we discuss our proposed scanning approach with advantages and disadvantages in Section 6. We discuss ethical considerations in Section 7 and conclude our paper in Section 8.

2 RELATED WORK

IPv4 Tarpits: Tools such as ZMap [13] and MASSCAN [19] made it feasible to conduct Internet-wide port scans. However, it is crucial to exercise caution when analyzing their output results, particularly in terms of the reachable services they may have identified. In 2014, Alt et al. [1] analyzed so-called tarpits, cyber-defense tools which imitate many fake hosts to slow down scanners. As tarpits are often deployed on a whole prefix, they look like HRPs on the Internet. Moreover, tarpits behave like responsive hosts but do not offer any real service. Therefore, our evaluation of HRP application-layer responsiveness shows that tarpits play only a minor role in our dataset.

We do not focus on this single reason for HRPs, but quantify their occurrence and specifics in general. 1.9 M IPv4 addresses were also assumed to be tarpits by Bano et al. [2]. Their argument rested on the responsiveness on a high port as such a port is unlikely to be responsive on many consecutive addresses. Similarly, we show that some HRPs are responsive on all ports, including high ports unrelated to any service. In 2021, Izhikevich et al. [23] investigated the phenomenon of hosts responding to SYN scans but not following through with the application-layer handshake performing probes to 0.1 % of the IPv4 address space. The authors built a tool to identify services running on unexpected ports, showing that many common services are in practice deployed on ports other than the default or well-known ones. They identified full IPv4 blocks larger than /24

with a zero TCP initial window but did not further investigate on prefix level responsiveness. In Section 5, we analyze the presence of HRPs in the dataset of Izhikevich et al. and validate our claims using their data.

Predicting Responsiveness: A more recent study by Izhikevich et al. [24] proposed an approach to predict service availability on non-standardized alternative ports. They perform small scale scans in order to deduce availability of addresses on the Internet. In this work, we show that the alternative and non-standardized ports suffer the most bias through HRPs. Although Izhikevich et al. verify service availability using application layer handshakes, we argue that considering our findings can improve the success rate and performance of such measurement campaigns.

Additional work by Durumeric et al. [11] introduced the Censys search engine. Our findings also extend to search engines such as Censys and the data they provide to researchers. ZMap has been employed for various purposes, including determining the number of vulnerable Heartbleed hosts [12] and tracking patching progress. Furthermore, Costin et al. [8] employed ZMap in the intended manner, identifying hosts with open TCP/443 ports for conducting HTTPS probes. Throughout this paper, we will demonstrate why it has become increasingly crucial to account for the presence of HRPs when conducting such evaluations.

Ethical Scanning: Klick et al. [25] introduced a scanning approach that reduces probing overhead by focusing on interesting, densely populated prefixes based on historical data. We confirm that HRPs are often stable and large fractions of responsive addresses are within HRPs. However, focusing on those potentially introduces biases in scans towards CDNs and TCP proxies. Moreover, we show that the application-layer success rate is higher outside HRPs, their approach would potentially omit these non-HRP targets.

IPv6 Aliased Prefixes: Subtypes of HRPs have been regularly discussed in IPv6 research. During IPv6 scans, even a single HRP can already be too large to be practically scanned. Murdock et al. [28], Gasser et al. [16], and others [3, 26, 30] identified so-called aliased prefixes that impact the statistical evaluation of scans. They attributed the effect to a single host that aliased an entire prefix. The authors developed means to detect this based on randomly sampled addresses within each prefix. Gasser et al. [16] included such a detection in their ongoing hitlist service. However, in 2022, Zirngibl et al. [43] showed that aliased prefixes are not necessarily a single host, *e.g.*, in CDNs. We find HRPs announced by similar ASes in IPv4, and show their visibility for ports not investigated by Zirngibl et al..

3 DATASETS

In this section, we describe the data sources we rely on for our analysis later in Section 5. Our sources specifically include port, HTTPS and DNS scans that we conduct ourselves, as well as (semi)public data sources that we use as complement. We discuss ethical considerations for our scans in Section 7. The datasets we obtain from the various sources mainly cover the first week of August 2022. We will note where this does not apply.

3.1 Port Scans

We conducted different port scans from three vantage points. Two vantage points (Munich and Saarbrücken) are located in Germany, while the third is located in Australia (Sydney). The first two vantage points are used for most of the scans. The third was used in sync with Munich to evaluate the visibility of HRPs from two geographically and topologically distant vantage points (see Section 5.3). We use ZMap and its TCP SYN scan module to perform these scans. The used scan module reports a target as being responsive, if a TCP SYN-ACK packet has been received from that target.

We run two types of scans from Munich and Saarbrücken:

(i) Long-running scan campaign for TCP port 443 on the announced IPv4 address space (Jan 2021 — Jan 2023). We use the resulting data to analyze the stability of HRPs over time.

(ii) A wider-range IPv4 scanning campaign targeting 36 different TCP ports (Aug 2022 – Sep 2022). We complement this scan data using port scan data from Rapid7's Project Sonar [38]. Our scans allow us to evaluate HRPs while considering potential artifacts on our scan machines, of the used blocklist, and scanned targets.

3.2 HTTPS Scans

Besides port scans, we run weekly HTTPS scans on responsive hosts obtained through the port scans. These application layer scans are conducted from Munich using the Goscanner [18], with and without Server Name Indication (SNI). The weekly TCP/443 port scans are used as the target list for the scan without SNI. The SNI scan uses a locally resolved A records dataset for more than 350 M domain names.

3.3 DNS Data

We use DNS data provided by the OpenINTEL project, which actively queries a significant part of all globally registered domain names on a daily basis [39]. The OpenINTEL measurement is primarily seeded using zone files, including well over 1 k zone files from ICANN's CZDS [22] as well as country-code top-level domain zones. The measurement data include address resolutions for domain names as well as for infrastructure records (i.e., MX and NS). The associated A records allows us to map domain hosting and mail exchangers and authoritative name servers to HRPs.

3.4 Rapid7 Data

To complement our own scans, we use data provided by Rapid7's Project Sonar [38]. Rapid7 provides measurement data from Internet-wide surveys to gain insights into global exposure to common vulnerabilities containing regularly-conducted port scans on 129 different TCP ports (*e.g.*, 80, 443, 25, 60000) and 19 UDP ports, (*e.g.*, port 53). Note that common ports (*e.g.*, 80 or 443) are scanned more frequently by Rapid7. We use these frequent scans to determine the stability of HRPs besides our results. Rapid7 also provides application-layer scan data. We rely on Rapid7 data to reduce the need to run our own measurements and to reach a large coverage of different ports and protocols.

We use different datasets provided by Rapid7:

(i) For our analysis, we use port scans conducted by Rapid7 from February 1, 2021 to the end of 2022.

(ii) We use TLS scans to evaluate HRPs in more detail. Rapid7 performs scans for TLS ports other than TCP/443, so we do not have to run these ourselves. We use the TLS handshake status and the collected certificate from scans in the first week of August 2022.

(iii) We also use application-layer HTTP scans for port 80 and alternative HTTP ports from Rapid7. These contain the complete raw HTTP responses received by Rapid7.

4 HRP DETECTION APPROACH

This section details how we collect, aggregate, and enrich reachability data per prefix and finally identify a responsiveness threshold for HRPs.

4.1 Data Collection and Aggregation

Our tool takes as input a list of IPv4 addresses that are reachable on a specific port, *e.g.*, from a ZMap scan. Therefore, HRPs are specific to a protocol and a corresponding port. In Section 5.1, we evaluate cross-port responsiveness. As a first step, we aggregate these input addresses into prefixes. A prefix length of 24 bits is a de facto limit for the most-specific, globally routable prefix length in
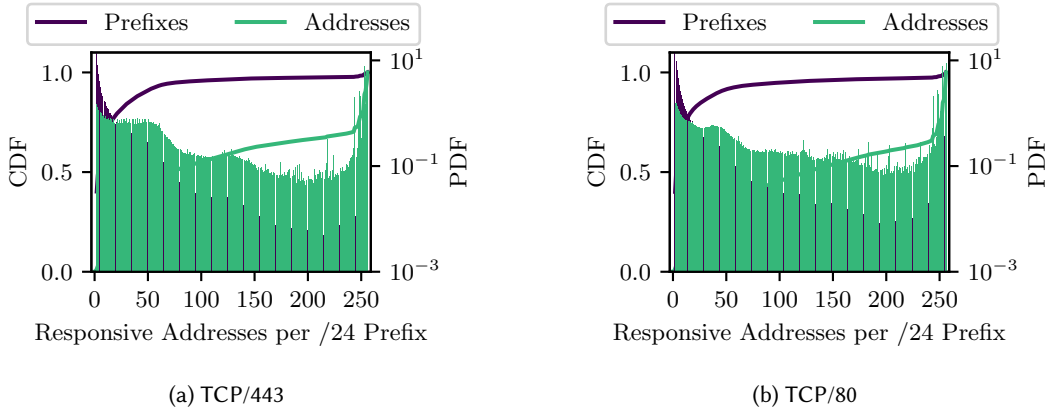
(a) TCP/443

(b) TCP/80

Fig. 1. Respective (bars – right y-axis) and cumulative (line – left y-axis) probability distribution of responsive addresses inside prefixes. The address data represents the influence of prefixes on the scan results.The prefix indicates that the majority of prefixes do not meet the criteria for HRPs. Note the logarithmic axis for the right Y-axis. More ports can be found on our published website [34] and in the appendix Figure 8.

IPv4 [10, 27, 35]. Thus, this constitutes the smallest IPv4 prefix that reliably propagates in the Border Gateway Protocol (BGP), representing the most limited subset of addresses we evaluate. Figure 2 shows that more-specific prefix sizes do not amount to a substantial share of high responsiveness in our analysis. In Section 5.1, we show that our results indeed support this choice.

When aggregating responsive addresses into /24 prefixes, we enrich each entry with additional contextual information, such as the origin AS number and the corresponding BGP prefix, using BGP dumps from a Route Views [29] collector of the respective scan date. Our approach is as efficient as the aggregation efficiency as we read the input data only once and linearly after aggregating it. This processing strategy enables parallelized (*e.g.*, sharded) processing.

4.2 Identifying a Threshold for Highly Responsive Prefixes

With the prefix statistics at hand, we next need to identify a suitable threshold for HRPs. Previous work mentioned the existence of fully responsive prefixes in IPv4 [23]. However, to the best of our knowledge, there has been no attempt yet to identify the various categories that together constitute HRPs. The need for a generalized definition that covers prefixes with similar behaviors arises from several aspects that bear relevance for scanning: *(i)* Packet loss can occur, causing a prefix to appear as not 100 % responsive, even when it actually is. It is important to note that even a minimal loss rate can substantially impact an Internet-wide IPv4 scan. *(ii)* Scalable port scanners such as ZMap sometimes trigger rate limiting on the receiver side. While built-in target randomization helps mitigate the risk of overburdening target systems, it cannot completely eliminate this problem. We also identify some instances of rate limiting and evaluate them in Section 5.3.

To obtain a viable and robust responsiveness threshold, we analyze the distribution of responsive addresses inside a prefix. In Figure 1a, we show a very relevant pattern for TCP/443 scans. More than 75 % of prefixes contain fewer than 15 responsive addresses, and 91 % include at most 50 addresses. However, there is a substantial increase in prefixes with 231 or more responsive addresses (*i.e.*, 90 % of a /24 prefix). Although these account for only 2.2 % of *visible /24 prefixes*, they account for *30 % of responsive addresses*. A similar pattern is visible for TCP/80 in Figure 1b. A comprehensive analysis encompassing all evaluated ports (see Figure 2) reveals a consistent pattern among all of
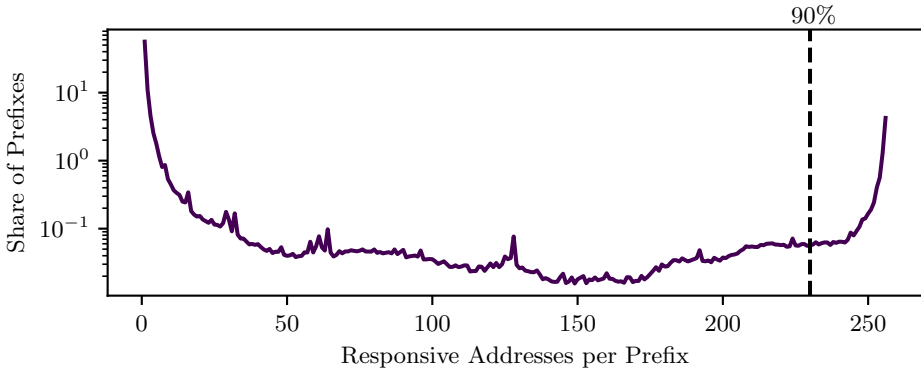
Fig. 2. Average probability distribution of responsive addresses inside prefixes for all prefixes. Note the logarithmic Y-axis. The vertical line marks the 90 % threshold.

them. Furthermore, separate evaluations on other ports are available in the appendix (see Figure 8) and on our website [34]. These supplementary evaluations also exhibit a similar pattern to what is observed for TCP/80, TCP/443, and the aggregated analysis. Consequently, this observation holds true not only for specific ports but also extends to all other ports.

To ascertain enough generality, we inspect the curve's knee and identify 90 % of responsive addresses as a suitable threshold. In Section 5, we compare our choice to a 95 % threshold and show that this leads to only marginal differences in the outcome of our investigation.

5 EVALUATION OF PORT SCAN HRPS

In this section, we evaluate the presence, stability, and origin of HRPs. Furthermore, we analyze whether application-layer services actually run in these prefixes. We use these insights to identify different classes of HRPs.

5.1 Comparison of Ports and Protocols

We apply our definition of HRPs from Section 4.2 to all available port scan data to obtain a more complete overview. In total, we analyzed scan data for 142 TCP and 19 UDP ports. Due to space constraints, we report only on well-known or particularly interesting ports here. Evaluations for all ports are available on our website [34].

Figure 3a shows the share of addresses in HRPs, measured against the full set of IP addresses that respond on a given port. The figure shows data from our scans (see Section 3). Except for the SSH port (TCP/22), at least 30 % of the responsive addresses are located within HRPs. The SSH port is considered sensitive, which may be why operational configurations treat it differently. Well-known ports (*e.g.*, TCP/80, TCP/443) and their alternative ports (TCP/8080 and TCP/8443) have an HRP address share of 30 %-40 %. Ports for less popular services and other alternative ports generally have a higher share of HRP addresses, which goes up to 80 %. In general, the smaller the overall set of responsive addresses and the less known a port is, the higher the fraction of HRP addresses will be. One reason for this is a group of HRPs that are responsive on all scanned ports (see Figure 4).

Figure 3b shows our results for UDP ports, taken from Rapid7 scan data. Compared to the TCP scans, the absolute responsiveness is drastically smaller by at least one order of magnitude. UDP scans generally require valid payloads to determine whether a port is truly responsive. However,

(a) Selected TCP ports from our scan.
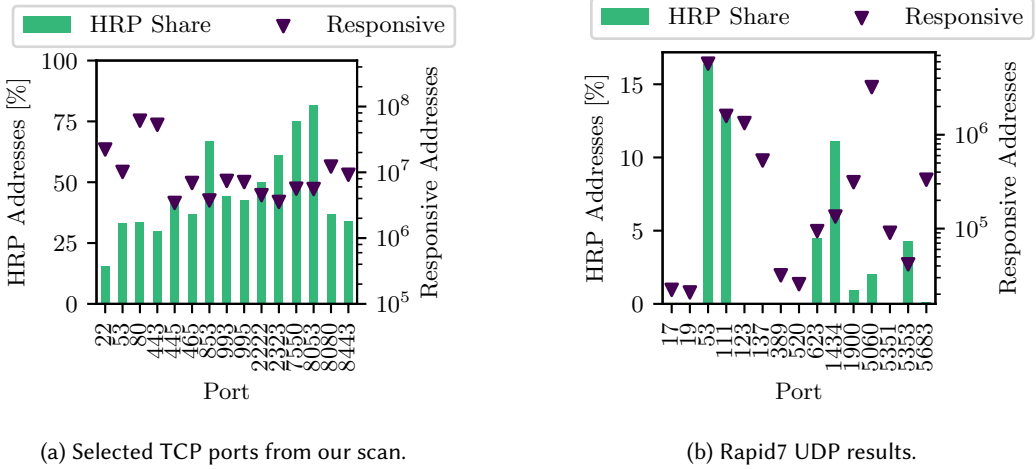
(b) Rapid7 UDP results.

Fig. 3. HRP address share and total number of responsive addresses per port. Note the logarithmic right Y-axis and the different Y-axis scaling for TCP and UDP.

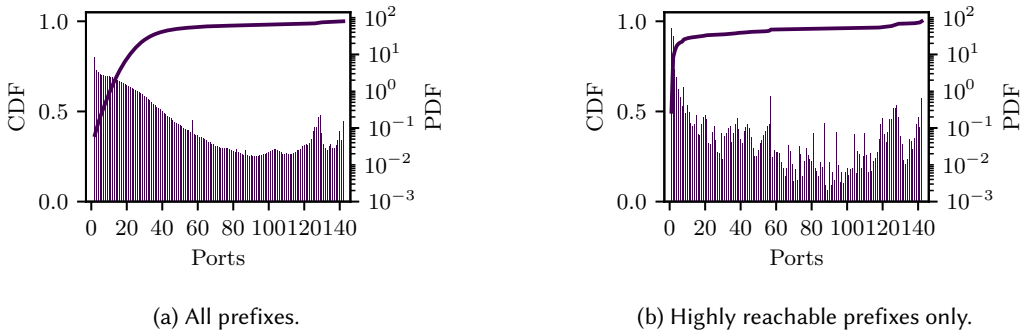(a) All prefixes.

(b) Highly reachable prefixes only.

Fig. 4. Number of ports on which addresses within a /24 prefix respond. The HRP analysis counts only ports which appear to be highly responsive for the prefix. Note the logarithmic right Y-axis.

the fraction of addresses within HRPs is smaller. We still find some HRPs, especially for more well known ports, *e.g.*, 53.

To get further insights into the specifics of HRPs, we analyze whether they are responsive for a single port, specific port combinations, or all ports in our dataset. Figure 4a shows the number of ports where at least one address is responsive inside the prefix. The same but limited to HRPs is plotted in Figure 4b. The peak at 129 ports in both figures is because Rapid7 scans only 129 TCP ports, and we merged our port scanning set with Rapid7. Some HRPs appear not responsive in our scans (see Section 5.3). Moreover, fewer prefixes are highly responsive on a large share of scanned ports in Rapid7 data. While our scans find 31 % of detected HRPs being classified as highly responsive for ten or more ports, Rapid7's data only accumulates to 13.6 % (our scans also accumulate in absolute numbers to more than double the Rapid7's amount).

We note the more pronounced values for HRPs: 80 % appear only for three ports or fewer. 50 % of all distinct HRPs are only classified as such based on the responsiveness of a single port. For port TCP/80, we find 59.8 k prefixes; for port TCP/443, we find 40.7 k; and for port TCP/25, we find
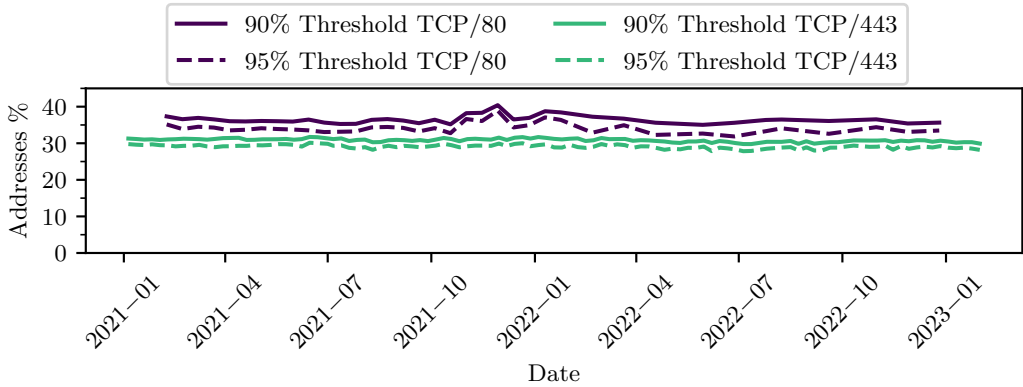
Fig. 5. Stability of highly responsive prefixes for weekly scans.

33.8 k. 30 % of prefixes are classified as HRP based on two responsive ports 80, and 443. However, 4700 (1 %) HRPs are responsive on at least 139 ports within our data.

Key take-away: *We find HRPs for all evaluated TCP ports and show that they cover between 30 % and up to 80 % of all addresses responsive on a given port. While 80 % of HRPs are classified as such based on one or two responsive ports, 5 % are responsive on more than 120 ports. Due to the nature of UDP scanning, the role of HRPs is less prominent. However, we can still find instances of UDP HRPs.*

5.2 Stability over Time

In this section, we analyze our long-running scans on TCP/443 together with scans of TCP/80 by Rapid7 to investigate the stability of HRPs over time. Apart from understanding the long-term behavior of HRPs, this also provides evidence that the analyses of the data collected for this paper do not depend on the time a scan was carried out.

Figure 5 shows the fraction of addresses within HRPs for each weekly port scan from Munich. We also compare our HRP definition, which is based on 90 % responsiveness, with an alternative definition based on 95 % responsiveness. We show that the difference is minimal, even over time. The data from Rapid7 shows slightly more variance, but the difference is at most 5 %. These numbers imply no significant increase in address responsiveness within each HRP. We obtained similar findings for other services and alternative HTTPS ports.

Turning our attention to the stability of individual HRPs, we observe that most prefixes exhibit a high degree of stability, although some churn is evident. During each TCP port 443 scan, an average of 65.4 k HRPs are identified, amounting to 110 k distinct HRPs over the two-year observation period. 78 k of these remain consistently visible and classified as HRPs for at least half of our observation period, and 38.3 % of these maintain their classification throughout the entire observation period. 73.1 % of prefixes that are normally classified as HRPs are missing this classification for at most five scans. It is plausible that these misclassifications are due to scanning artifacts, *e.g.*, because of rate limiting on the receiver's side (see Section 5.3). The prevalence of HRPs that we have identified, both spatially and over time, in this and the preceding section underscores the importance for researchers to account for their influence when analyzing scan data.

Key take-away: *The contribution of HRPs to the overall share of responsive addresses remains consistent over a span of two years. Over a year, more than 73 % of highly responsive prefixes remain consistently*
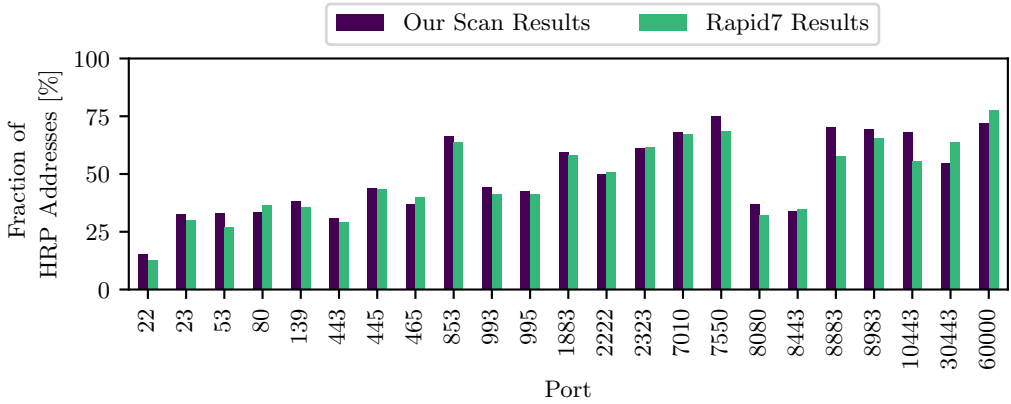
Fig. 6. Comparing HRP shares of Rapid7 data and our own scanning results.

detectable in our scans. Consequently, it is crucial to account for HRPs in the analysis of both historical and recently collected scan data.

5.3 Spatial Stability

In order to determine the influence of the vantage point on the detection of HRPs, we run different measurements and comparisons between geographically and topologically diverse vantage points. Figure 6 compares our port scans from August 2022 with Rapid7's results for 23 ports from the same time frame. While we have observed minor variations, there are no significant discrepancies between our results and those obtained by Rapid7.

To gain deeper insights into why such discrepancies arise, we conducted parallel scans from Munich (Europe) and Sydney (Oceania) employing the same ZMap parameters. This approach ensures a fully comparable scan result as all hosts are scanned at about the same time and from different vantage points. We use TCP/443 and TCP/55974 as the destination ports, the latter being a random ephemeral port to find differences in HRP deployments on all ports.

We have determined that 1 % of HRPs identified at one vantage point are not classified as HRPs at another. The majority of these unclassified prefixes have no responsive hosts. Upon conducting additional verification scans from Munich on prefixes that were exclusively detected by Sydney, we found that 18 % of these prefixes displayed responsiveness in the second scan and were accurately categorized as HRP. The responsive prefixes are part of the Cloudflare AS (AS13335). Sydney has a similar number of Cloudflare prefixes not visible in its scan but rated as HRP by Munich. Therefore, we conclude this to be an artifact of rate limiting on Cloudflare's side. The remaining prefixes, which are only seen by Sydney, are still unresponsive to the Munich scan. We assume this to be blocking of scans from Munich, possibly due to previous, long-running measurement campaigns from the same /24 subnet.

TCP/55974 results show a similar number of differences with three ASes (DoD: AS721 and AS5972, Illinois Century Network: AS6325) contributing 96 % of the missing HRPs in the Munich results. As DoD accounts for 11 % of visible prefixes and 39 % of responsive hosts, we also attribute this effect to rate limiting.

Similar insights on the impact of vantage points have been reported by Wan et al. [40] in 2020. Given that Rapid7 may not be utilizing precisely identical parameters, conducting scans from a distinct network (one that hosts several Internet-wide scanning campaigns), and is scanning hosts

Table 1. Top 10 ASes based on the number of HRPs we detect across all scans in August 2022. The HRP share is the degree to which an AS is filled with HRPs.

AS		Visible /24	HRPs	HRP Share	Visible Ports	Ports with HRPs
Akamai	AS16625	22.9k	22.4k	97.8%	5	3
Akamai	AS20940	24.7k	21.1k	85.6%	136	5
Telin	AS7713	12.5k	6.5k	52.5%	136	4
Amazon	AS16509	134.9k	6.0k	4.4%	136	135
DoD	AS721	4.9k	4.5k	91.3%	136	55
DoD	AS5972	4.5k	4.4k	99.3%	60	54
du	AS15802	4.3k	4.1k	96.0%	136	3
Cloudflare	AS13335	3.1k	3.0k	98.3%	136	136
Cogent	AS174	17.0k	2.8k	16.5%	136	136
TANet	AS1659	8.2k	2.4k	29.8%	136	135

at different points in time, the discrepancies between our results and those of Rapid7 are minimal. These variations are expected and align with the outcomes of our controlled experiments.

Key take-away: *We find differences in the order of 1 % between different vantage points when applying the same scanning procedure. These results confirm previous results by Wan et al. [40]. Well-known ports have comparably fewer differences, primarily due to their higher responsiveness. The majority of discrepancies can be attributed to rate limiting. Consequently, we conclude that it is feasible to utilize detected HRP prefixes from different vantage points.*

5.4 Origin ASes

The HRPs we find are announced by 67.2 k origin ASes. We determine the origin AS with BGP dumps from RouteViews [29] from the day of the respective scan. Next, we investigate which ASes originate HRPs and to what degree an AS may deploy features that make a prefix an HRP.

Of the ASes that originate HRPs, 42 % have HRPs for a single port; 76 % of ASes have HRPs with at most five different, responsive ports. 118 ASes announce at least one HRP where the IP addresses respond on all ports. Table 1 shows the top ten ASes with at least one HRP. Cloudflare (AS13335), for example, announces a /16 that is fully responsive on all ports. We identified that Spectrum [4, 6, 7], a reverse TCP/UDP proxy for DDoS protection, is related to the affected range. We contacted Cloudflare and confirmed with them that the service is built on the principle of binding to all ports. Google, the second largest AS with an HRP for all ports, offers a similar service.

In Table 1, we find Akamai at the top of the list with 22 k HRPs in AS16625 and 21 k in AS20940. The latter AS has nearly all of its prefixes where we find responsive addresses also classified as HRPs. Notable other ASes are Amazon, which has 6 k HRPs, but these cover only 4.4 % of its reachable prefixes. The ASes with an HRP on all ports are all large network infrastructure providers: Cloudflare, Amazon, Google, Cogent, and TANet[1]. It is not too surprising that CDNs are commonly connected to HRPs: they deploy various techniques [14] to support millions of domains served on limited address space. ASes primarily providing CDN services consist of up to 99 % of HRPs compared to their overall announced address space (see HRP share in Table 1). However, this impact on scan data has so far received little attention.

We also evaluate the AS distribution per port. For port TCP/443, Akamai has the most HRPs, followed by Google, Cloudflare, Amazon, and Fastly. These five organizations cover 64 % of all port 443 HRPs. While Akamai's AS20940 has more port 443 HRPs, AS16625 (also Akamai) has more

---

[1]Amazon and TANet are only missing out on a single port. See Section 5.3 for more information.

than double the amount of HRPs on port 80 compared to port 443. This result may reflect different use cases that the prefixes inside these ASes serve.

CDNs affinity to HRPs: We identified several CDN ASes in our results that are nearly entirely populated with HRPs. For instance, Akamai exhibits 97.8 % and 85.6 % of announced /24 prefixes classified as highly responsive for two different ASes, while Cloudflare demonstrates 98.3 %, and Fastly registers 87.6 %. Therefore, we set out to uncover the reasons for CDNs using HRPs. For Cloudflare, we find a paper from Fayed et al. [14] with an accompanying blog post [5] that explains Cloudflare's *addressing agility* approach. This technique decouples IP addresses from domain names and services. The authoritative name server can select the addresses in the query response from a full prefix. Hence, all addresses inside this prefix have to be responsive, and Cloudflare's approach needs to handle all prefix-assigned services. We verified this assumption by registering a test website with Cloudflare, resolving the domain to an A record, and connecting to several addresses inside the A record's /24 prefix. Each TLS handshake, which included our registered domain name in the server name indication extension to any address inside this prefix, resulted in the same certificate for our domain name.

Key take-away: *In this section we show the that the distribution across ASes is dominated by CDNs and other content providing ASes. HRPs make up a substantial part of these ASes. We could confirm that Cloudflare's address agility technique is responsible for Cloudflare's large share of HRPs. We assume that other CDNs deploy similar techniques.*

5.5 Application-Layer Results

A number of published papers [2, 23, 43] have reported that an open port does not necessarily imply that the corresponding application is going to respond as well. We use application-layer scans to reveal whether the services with open ports in the respective HRPs actually respond. Izhikevich et al. [23] showed that the services running on given ports can vary widely, especially for ports that are not among the well-known ones. Therefore, our analysis focuses on well-known ports for HTTP/HTTPS, email, and DNS services. We also include commonly used HTTPS alternative ports. We use HTTPS scans from Munich, TLS scans for HTTPS alternative ports (including email and DNS ports) from Rapid7, and HTTP GET requests from Rapid7[2].

Table 2 shows the results of application-layer scans for targets within our detected HRPs. We differentiate between *(i)* the availability of the expected application-layer service on at least one address and *(ii)* the availability of the service on more than 90 % of all previously responsive addresses.

For 86 % of TCP/80 and 84 % of TCP/443 HRPs, we find at least one host with a successful application-layer handshake. CDNs play a major role as the top five organizations cover 64 % of HRPs. For 27.7 k of TCP/443 HRPs, or 49 % of application-layer responsive HRPs, more than 90 % of the respective addresses inside each prefix are responsive. For the latter category, the fraction of HRPs with HTTP on TCP/80 is substantially higher (81 %) than that for TLS on TCP/443. We note, however, that scanning a port with a TLS scanner can result in a specific TLS error, *e.g.*, when a SNI is not given (as is the case with the scans by Rapid7), even though the corresponding hosts still offer the expected application-layer service in principle. When disregarding the targets with such errors[3], the share of > 90 % application-layer service HRPs in TCP/443 is slightly higher (89 %) than that for TCP/80.

---

[2]Note that Rapid7 TLS scans are performed without SNI while our TLS scans from Munich include SNIs.
[3]We omit these cases from Table 2 in favor of successful connections to provide useful references for the single identifier evaluation.

Table 2. Application-layer responsiveness for detected HRPs. For TLS we use the certificate as identifier. For HTTP a hash of the response data.

Port	# HRPs	App. Layer Success		Same Identifier	
		# HRPs	>90 % Success	# HRPs	HRP [%]
80	91 674	79 234	64 539	760	0.01
443	64 435	54 203	26 715	2718	0.1
1443	14 461	557	364	33	9.1
4443	9722	728	474	141	29.7
8443	13 048	3287	809	384	47.5
25	33 294	3493	2210	2041	92.4
110	11 394	2553	2379	1944	81.7
143	11 112	2727	2527	2082	82.4
465	9076	2793	2627	2174	82.8
587	11 274	2067	1056	931	88.2
993	10 362	2906	2703	2266	83.8
995	10 717	2910	2736	2303	84.2
853	8352	565	379	53	14.0

We evaluate the application-layer scans to determine if we find different answers from IP addresses inside an HRP. For TLS, we use the certificate as an indicator (we do not collect HTML responses at Munich). Different certificates may indicate that these HRPs are unlikely to be aliased prefixes. However, different answers are not necessarily indicators of the presence of distinct hosts, either [1]. To compare HTTP responses of different targets, we use a hash of the returned HTML responses. This is more error-prone due to session-specific values such as the HTTP Date header or cookies. Therefore, the results for HTTP must be interpreted as a lower bound.

Table 2 shows that only 0.01 % of HRPs for port 80 and 0.1 % of HRPs for port 443 use the same application-layer identifier within the prefix. This is expected, as many CDNs offer HTTPS services, which are unlikely to return unique certificates, especially when scanned with SNI [15]. In contrast, the TLS scans for email ports reveal that 80 % of HRPs serve a single certificate on all scanned addresses. *home.pl*, a large polish web and mail hoster, is responsible for more than 30 % of such HRPs. We also find other similar, smaller providers within this group.

In comparison to the officially assigned HTTPS port 443, the alternative ports have a much smaller fraction of HRPs providing an application-layer service. Only 4 %–25 % of HRPs respond successfully on at least one address. We also find a substantially higher proportion of identical certificates for alternative HTTPS ports in the "> 90 % successful" category.

We also compare application-layer success of scanned targets within and outside HRPs in Table 3.It clearly shows that our scans succeed more often for addresses not located within an HRP. In the case of TCP/80 and TCP/443 we find about 89 % of addresses outside HRPs serving the expected service, while inside HRPs it is only 61 % for TCP/443 and 75 % for TCP/80. For addresses inside HRPs we analyze the subset of prefixes where more than 90 % of addresses expose the expected service. This subset is responsible for more than half of all successful responses. Digging further in this subset we find that ports related to mail services often seem to use the same certificate. Therefore, a scan covering all responsive addresses not only often fails for HRP addresses, but when a connection can actually be established the information obtained is almost entirely redundant. This data also reflects our results from Section 5.4, where we find two causes for HRPs: TCP proxies and CDN deployments of Web services.

Table 3. Comparison of successful application-layer connections between non-HRP and HRP addresses. The last two columns represent the relative shares compared to the preceding column, with their subset (App. Layer > 90 %) containing addresses within HRPs meeting the 90 % successful connections criterion. Note: This data is on addresses while Table 2 is on prefixes.

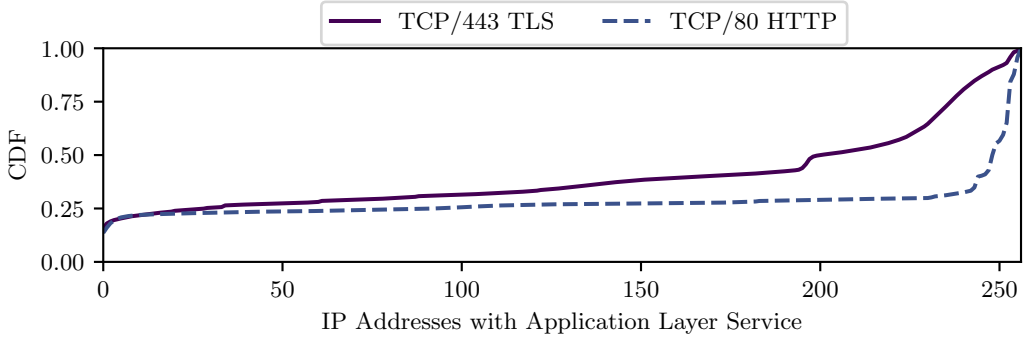| | | | App. Layer > 90 % | |
Port	Non-HRP	HRP	HRP Share	Same Identifier
80	89.1%	73.5%	95.4%	1.2%
443	88.3%	61.1%	64.9%	10.6%
25	32.7%	6.7%	97.9%	92.3%
110	59.8%	20.8%	99.2%	81.9%
143	61.3%	22.6%	99.5%	82.6%
465	65.9%	28.7%	99.5%	82.9%
587	34.9%	12.7%	73.2%	87.4%
993	70.9%	26.1%	99.1%	83.7%
995	69.4%	25.4%	99.4%	84.1%
1443	7.0%	2.8%	88.4%	9.1%
4443	24.2%	5.3%	90.6%	29.7%
8443	67.7%	11.0%	55.2%	46.8%

Fig. 7. Cumulative distribution of IP addresses providing a TLS or HTTP service within HRPs.

Figure 7 shows the cumulative distribution function (CDF) of HRPs with successful application-layer scans for ports 80 and 443. Both protocols have close to 20 % HRPs with unsuccessful application handshakes. While TLS has 41 % of HRPs with more than 90 % availability, more than 82 % of HTTP HRPs have at most 40 successful application-layer handshakes. We also find 16 % of application-layer responsive HRPs with more than 90 % successful HTTP connection attempts. Other TLS ports show a distribution similar to that for HTTP, with the difference that > 80 % of HRPs do not respond on a single address (see also Table 2).

Finally, we also analyzed published data by Izhikevich et al. [23]. They provide the dataset where they scanned a sample of the IPv4 address space. In total, the published dataset contains data on 142 M port scans for 307 k IP addresses. We compare these scans with our HRPs on all available ports and locations from August 2022. As our dataset contains only data on 142 ports we approximate if a prefix could be highly responsive on the scanned port by checking if it was classified as HRP at a location for ten or more ports (see Section 5.1). 50 M LZR scans are within

Table 4. DNS statistics derived from NS records for DNS services and MX records for mail services with names which resolve to an IP address inside HRPs. The SLDs column represents the number of SLDs using such NS/MX records.

	HRP	IP addresses		FQDNs	SLDs
TCP/53	12.0%	40.9k	1.4%	161.6k	115.6M
UDP/53	25.5%	29.0k	3.1%	133.0k	104.6M
TCP/853	2.3%	1.0k	0.1%	1.2k	55.2k
Mail Ports:					
TCP/25	18.5%	172.4k	2.0%	3.0M	3.7M
TCP/110	26.4%	126.0k	4.4%	2.7M	3.2M
TCP/143	26.3%	121.6k	4.3%	2.7M	3.2M
TCP/465	30.4%	121.0k	5.3%	2.7M	3.2M
TCP/587	30.6%	133.1k	4.7%	2.7M	3.2M
TCP/993	27.2%	120.0k	4.6%	2.7M	3.2M
TCP/995	27.5%	261.5k	4.6%	5.6M	6.6M

this threshold and a comparison of scans within and outside this set reveals a significantly better response rate (63 % vs 91 %) for scans outside HRPs.

Key take-away: *When analyzing application-layer responsiveness, we observe a relatively higher share of responsive hosts outside HRPs as opposed to those within. The application-layer analysis of deployment-specific properties reveals that alternative Web ports and mail ports often share a single identifier across an HRP. Finally, our validation on external scanning data reaffirms that application-layer scans performed on targets outside HRPs yield a higher success rate compared to targets within.*

5.6 DNS Analysis

We use DNS data provided by the OpenINTEL project to evaluate the theoretical population and addressing diversity of HRPs. As we evaluate DNS data, we note that the information provided in DNS records may often come from a party other than the owner of a HRP; this is quite different from active scans into HRPs, which establish ground truth. Theoretically, anyone can create DNS A records pointing to a specific address. However, it is plausible to assume most DNS records express specific operational choices and may thus help to reveal the nature of an HRP. We use DNS results in our HTTPS scans and discuss its effect in Section 6.

Table 4 shows how many fully qualified domain names (FQDNs) and Second Level Domains (SLDs) are associated with specific ports and protocols (DNS and email) on average. We consider the respective record types (*i.e.*, NS for DNS and MX for email) and map a name server or mail server to HRPs if at least one of its A records points to an address within the HRPs. For mail services we assume a host used for MX records also provides POP3 or IMAP4 services. The table also shows the number of unique IP addresses contained in A records, and how many domain names reference the associated name servers and mail exchangers.

UDP/53 HRPs are usually referred to by DNS NS records. While only about 3 % of possible HRP addresses are actually present in name server resolutions, these name servers serve more than 100 M SLDs. The smaller results for TCP/853 are expected as DNS-over-TLS is not very widely deployed and is more commonly used by resolvers, not nameservers. If we consider the IP address space covered by DNS HRPs, we observe that some HRP are referenced on at least 90% of their addresses by some NS record. Specifically, this involves 28 TCP/53 and 15 UDP/53 HRPs. Although an open port is not necessarily indicative of service availability, DNS references to HRP addresses as well as open DNS ports therein suggest that DNS services are present in the respective HRP.

Table 5. Unique number of IP addresses, the fraction of HRP address space they cover, and FQDNs derived from A records for HTTP/HTTPS ports.

	HRP	IP addresses		FQDNs
TCP/80	34.4%	4.7M	11.0%	171.4M
TCP/443	30.8%	2.0M	6.3%	149.1M
TCP/8000	32.7%	349.2k	10.9%	2.1M
TCP/8080	45.9%	504.2k	12.6%	27.3M
TCP/8443	56.8%	517.3k	16.7%	28.1M

We find more MX records that point into HRPs than NS records. There are also significantly more unique MX FQDNs than NS FQDNs. However, the number of SLDs related to MX records is only about the same number as the distinct MX record names. Depending on the port, 1% to 2% of HRPs have at least 90% of their respective addresses referenced via MX records.

We derive possible use of HTTP/HTTPS from DNS data as well by inspecting the A records for domain name's apexes. Table 5 summarizes the results. Of the port-specific HTTP HRPs, 30%—57% are related to an A record in the DNS. Nearly half of all HRPs with popular alternative ports are referenced in the DNS data. Interestingly, fewer specific IP addresses in TCP/443 HRPs are referenced compared to TCP/80 HRPs.

A large CDN presence is also visible in the DNS data. 149 M FQDNs resolve to TCP/443 HRPs and 171 M to TCP/80 HRPs. With this in mind, it is not surprising to find Alexa Top 1M domains resolving to 366 k IP addresses inside TCP/443 HRPs and 377 k IP addresses inside TCP/80 HRPs. Therefore, DNS data shows that while only few IP addresses are referenced, these are used by many millions of FQDNs. Nevertheless, more than 80% of the address space is not referred to by DNS. Hence, domain-based measurement studies hit only a fraction of addresses compared to seeding the input from port scans.

Key take-away: *The analysis of DNS data provided by OpenINTEL shows that only a fraction of responsive addresses is actually referred to in DNS. Millions of FQDNs resolve to addresses inside HRPs or refer to MX and NS domains resolving to addresses inside HRPs. Therefore, HRPs can contain important hosts, but the number of referred addresses is smaller and can be identified via DNS resolutions.*

6 HRP-AWARE SCANNING TECHNIQUE

We evaluated the presence of HRPs and showed that HRP addresses can account for up to 80% of all responsive addresses, especially in the case of non-well-known ports. We provided evidence that HRP addresses often expose few or no application-layer services at all. Our findings lead us to conclude that many prefixes are of limited value for scans and can incur bias in the results. At the same time, our work confirms—and gives broader and firmer evidence—for a previous result, namely that simple port scans are insufficient to determine service availability and liveness. The conclusion is that while application-layer scans are necessary for the reliable assessment of service deployments, the approach to carrying these out needs to be optimized.

Internet-wide application-layer scans typically commence with a port scan (e.g., employing ZMap) to identify responsive addresses. Subsequently, an application-layer scan is executed, focusing solely on the IP addresses that have responded to the port scan. Application-layer scans are more time-consuming due to extended message exchanges and additional round-trips, and they also impose a heavier load on the scanned entities. Therefore, it becomes imperative to minimize the scanning scope to what is genuinely essential. One way to achieve this optimization is to consider HRPs before initiating the scan. Consequently, our proposal aims to align with this objective.

6.1 Approach

We suggest using a list of HRPs, as produced by our tool, and scan these prefixes strategically. The use case will determine the exact strategy, but the following methods can be used and combined. If available, researchers should use data from DNS resolutions to determine IP addresses where services might be deployed inside an HRP. Scanning based on domain names is important in a number of cases, but especially for TLS scans, where omitting the SNI may lead to inaccurate results. If a use case cannot be addressed by scanning based on resolved domain names, we suggest a hybrid approach: instead of scanning HRPs in full, we suggest taking a sample that consists of IP addresses that are associated to DNS records (A, NS, MX, etc.) plus addresses sampled uniformly from the remainder of the HRP. The selection of appropriate domain names, by using the SVCB/HTTPS DNS records [42] to find service names and excluding parked domains [41], can improve the obtained results. If, for example, we decide to take a sample of ten addresses per HRP, and we find only five different IP addresses in DNS, we suggest filling up the remaining five addresses through sampling.

The sampled data will yield results that can be used to make further decisions. We identify three scenarios: *(i)* the HRP seems to be a proxy, *(ii)* the HRP behaves like a CDN and always returns the same result (*e.g.*, the same TLS certificate), and *(iii)* different application-layer responses for different hosts are returned. In the first two scenarios, the research question must be considered to decide whether further samples are necessary for statistically strong results, or whether the sample data is already enough. The third scenario describes an HRP which we suggest scanning in full.

6.2 Evaluation

We evaluate the effect of this approach on our own TLS scans by differentiating between the SNI scan which uses DNS resolutions as input list and the non-SNI scan which uses plain IP addresses from the port scan. We find that 30 % of the SNI scan targets, which can be successfully scanned, are within an HRP, while 80 % of targets in the non-SNI scan are. In Section 5.6 we showed that DNS scans provide a high quality input list, but we find that these also omit responsive HRPs. Therefore, applying a sampling rate on HRPs which are not covered by DNS is highly beneficial. The sampling step also helps to prevent unnecessary application-layer handshakes in case of redundant responses. A trivial sampling algorithm which selects ten addresses uniformly from a /24 successfully completes connections to 92 % of HRPs compared to the full scan.

A similar evaluation of Rapid7 TLS data for TLS alternative ports and other TLS-enables services, serves to validate our observations. Rapid7's dataset is not as detailed as our own scans as they only provide us data of successful scans, and exclusively perform TLS scans without SNI. However, by approximating the number of scans based on their TCP port scans, typically conducted a few days prior to the TLS scans, we can derive the missing numbers considering they use the standard approach (ZMap followed by the application layer scan). Our proposed approach yields even more favorable results for these additional ports, as it still collects over 99 % of unique certificates while reducing application layer scans by up to 75 % depending on the specific ports being assessed (number based on evaluations in Section 5.1 and Section 5.5). These findings demonstrate the broad applicability and efficacy of our approach, not being limited to specific ports.

6.3 Approach in Practice

Our HRP filter tool needs the full port scan results to determine which prefixes are highly responsive. Therefore, one option is to run the port scan first, then apply our filter, and only afterwards starting the application-layer scan. One can also use a previous scan result or the weekly HRP results which we publish. Either will work as we have shown the stability of HRP in time (see Section 5.2) and space (see Section 5.3). Therefore, the approach also applicable to data sources such as Rapid7's

Project Sonar [38] and Censys [11] port scans. We note that HRPs can be filtered during scanning and hence would not interfere when the timing of application-layer scans is critical.

6.4 Pros and Cons

Our approach offers two substantial advantages. It has the potential to reduce scanning effort by 20 %-80 %, as determined through our HRP analysis. Especially many of the TLS alternative port HRP addresses exhibit a low success rate. Given that timeouts are typically the most time-consuming aspect during TLS handshakes, their exclusion provides a substantial benefit. Nevertheless, our approach is not without limitations. First and foremost, it cannot ensure the comprehensive scanning of all reachable, distinct systems. This limitation may pose challenges when the scanning objective is to identify every individual host with a particular configuration, as is often the case in certain vulnerability scans. However, it is essential to recognize that when the research goal centers around statistical aggregates, our approach contributes to reducing bias and is advantageous for the overall research outcome. Based on our results, we believe that many application-layer scanning campaigns can lower their impact on the network and improve their results by considering our approach.

7 ETHICAL CONSIDERATIONS

We apply strict ethical measures to all our scans [9] and adopt community best practices [31]. We uphold this ethical approach by imposing limitations on our scan rate and maintaining an internal blocklist shared by all vantage points to record opt-out requests. Consequently, if a party requests not to be scanned, all vantage points will cease scanning their systems. Furthermore, each scanning machine identifies itself through reverse DNS, a hosted website, and informative WHOIS entries. Besides conducting all scans ourselves, we verify that findings are similar from different vantage points (see Section 5.3). This enables us to leverage pre-existing data from Rapid7 for the majority of port and application-layer scans. Furthermore, scans conducted across multiple vantage points are less likely to trigger intrusion detection systems, thanks to the reduced load from a single scanning entity. Consequently, they are less disruptive for network operators. Similar measures are applied by Rapid7: They publish their scanning prefixes, have an opt-out process, and describe more relevant considerations on their website [20].

8 CONCLUSION

In this paper we developed and presented a technique to detect highly responsive prefix (HRP) and analyzed the presence of HRPs on 142 TCP and 19 UDP ports in the IPv4 Internet. We developed a tool to detect HRPs and used it to analyze port scans for 142 TCP ports. We found that HRPs are highly visible on any analyzed TCP port and can make up to 80 % of responsive addresse. CDNs are the largest entities exposing HRPs for well-known Web ports. In some cases, CDNs deploy reverse proxies, which results in HRPs responsivity on all available ports. We proposed different approaches towards more efficient scanning techniques, depending on the research goals in question and showed that these can significantly reduce the number of targets to be scanned. Finally, we released tooling to this end, and will publish data and statistics on HRPs going forward.

REFERENCES

[1] Lance Alt, Robert Beverly, and Alberto Dainotti. 2014. Uncovering Network Tarpits with Degreaser. In *Proceedings of the 30th Annual Computer Security Applications Conference* (New Orleans, Louisiana, USA).

[2] Shehar Bano, Philipp Richter, Mobin Javed, Srikanth Sundaresan, Zakir Durumeric, Steven J. Murdoch, Richard Mortier, and Vern Paxson. 2018. Scanning the Internet for Liveness. *ACM SIGCOMM Computer Communication Review* (2018).

[3] Robert Beverly, William Brinkmeyer, Matthew Luckie, and Justin P. Rohrer. 2013. IPv6 Alias Resolution via Induced Fragmentation. In *Proc. Passive and Active Measurement (PAM)*.

[4] Cloudflare. 2019. *It's crowded in here!* https://blog.cloudflare.com/its-crowded-in-here/

[5] Cloudflare. 2021. *Unbuckling the narrow waist of IP: Addressing Agility for Names and Web Services.* https://blog.cloudflare.com/addressing-agility/

[6] Cloudflare. 2023. *Cloudflare Spectrum.* https://www.cloudflare.com/products/cloudflare-spectrum/

[7] Cloudflare. 2023. *Cloudflare Spectrum - Network ports.* https://developers.cloudflare.com/fundamentals/get-started/reference/network-ports/

[8] Andrei Costin, Jonas Zaddach, Aurélien Francillon, and Davide Balzarotti. 2014. A Large-Scale Analysis of the Security of Embedded Firmwares. In *23rd USENIX Security Symposium (USENIX Security 14)*. USENIX Association, San Diego, CA, 95–110.

[9] David Dittrich, Erin Kenneally, et al. 2012. The Menlo Report: Ethical principles guiding information and communication technology research. *US Department of Homeland Security* (2012).

[10] J. Durand, I. Pepelnjak, and G. Doering. 2015. BGP Operations and Security. RFC 7454 (Best Current Practice). https://doi.org/10.17487/RFC7454

[11] Zakir Durumeric, David Adrian, Ariana Mirian, Michael Bailey, and J. Alex Halderman. 2015. A Search Engine Backed by Internet-Wide Scanning. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security* (Denver, Colorado, USA) *(CCS '15)*. Association for Computing Machinery, New York, NY, USA, 542–553.

[12] Zakir Durumeric, Frank Li, James Kasten, Johanna Amann, Jethro Beekman, Mathias Payer, Nicolas Weaver, David Adrian, Vern Paxson, Michael Bailey, and J. Alex Halderman. 2014. The Matter of Heartbleed. In *Proceedings of the 2014 Conference on Internet Measurement Conference* (Vancouver, BC, Canada) *(IMC '14)*. Association for Computing Machinery, New York, NY, USA, 475–488.

[13] Zakir Durumeric, Eric Wustrow, and J. Alex Halderman. 2013. ZMap: Fast Internet-wide Scanning and Its Security Applications. In *Proc. USENIX Security Symposium*. Washington, D.C., USA.

[14] Marwan Fayed, Lorenz Bauer, Vasileios Giotsas, Sami Kerola, Marek Majkowski, Pavel Odintsov, Jakub Sitnicki, Taejoong Chung, Dave Levin, Alan Mislove, Christopher A. Wood, and Nick Sullivan. 2021. The Ties That Un-Bind: Decoupling IP from Web Services and Sockets for Robust Addressing Agility at CDN-Scale. In *Proceedings of the 2021 ACM SIGCOMM 2021 Conference (SIGCOMM '21)*. Association for Computing Machinery, New York, NY, USA.

[15] Oliver Gasser, Benjamin Hof, Max Helm, Maciej Korczynski, Ralph Holz, and Georg Carle. 2018. In Log We Trust: Revealing Poor Security Practices with Certificate Transparency Logs and Internet Measurements. In *Passive and Active Measurement Conference 2018*.

[16] Oliver Gasser, Quirin Scheitle, Pawel Foremski, Qasim Lone, Maciej Korczynski, Stephen D. Strowes, Luuk Hendriks, and Georg Carle. 2018. Clusters in the Expanse: Understanding and Unbiasing IPv6 Hitlists. In *Proc. ACM Int. Measurement Conference (IMC)* (Boston, MA, USA).

[17] Oliver Gasser, Quirin Scheitle, Sebastian Gebhard, and Georg Carle. 2016. Scanning the IPv6 Internet: Towards a Comprehensive Hitlist. In *Proc. 8th Int. Workshop on Traffic Monitoring and Analysis*. Louvain-la-Neuve, Belgium.

[18] Oliver Gasser, Markus Sosnowski, Patrick Sattler, and Johannes Zirngibl. 2023. *Goscanner.* Retrieved 2023-03-24 from https://github.com/tumi8/goscanner

[19] Robert Graham. [n. d.]. *MASSCAN: Mass IP port scanner.* https://github.com/robertdavidgraham/masscan

[20] Marcia Hofmann. 2013. *Legal Considerations for Widespread Scanning.* Retrieved 2023-09-26 from https://www.rapid7.com/blog/post/2013/10/30/legal-considerations-for-widespread-scanning/

[21] https://csirt.divd.nl/. 2023. *Making the internet safer through Coordinated Vulnerability Disclosure.* Retrieved 2023-03-24 from https://csirt.divd.nl/

[22] ICANN. 2023. *CZDS - Centralized Zone Data Service.* Retrieved 2023-10-05 from https://czds.icann.org/

[23] Liz Izhikevich, Renata Teixeira, and Zakir Durumeric. 2021. LZR: Identifying Unexpected Internet Services. In *Proc. USENIX Security Symposium.* https://www.usenix.org/conference/usenixsecurity21/presentation/izhikevich

[24] Liz Izhikevich, Renata Teixeira, and Zakir Durumeric. 2022. Predicting IPv4 Services across All Ports. In *Proceedings of the ACM SIGCOMM 2022 Conference* (Amsterdam, Netherlands) *(SIGCOMM '22)*. Association for Computing Machinery, New York, NY, USA, 503–515. https://doi.org/10.1145/3544216.3544249

[25] Johannes Klick, Stephan Lau, Matthias Wählisch, and Volker Roth. 2016. Towards Better Internet Citizenship: Reducing the Footprint of Internet-Wide Scans by Topology Aware Prefix Selection. In *Proc. ACM Int. Measurement Conference (IMC)* (Santa Monica, California, USA). Association for Computing Machinery, New York, NY, USA.

[26] Matthew Luckie, Robert Beverly, William Brinkmeyer, and kc claffy. 2013. Speedtrap: Internet-Scale IPv6 Alias Resolution. In *Proc. ACM Int. Measurement Conference (IMC)* (Barcelona, Spain).

[27] MANRS. 2021. Prefix filter configuration tools. https://www.manrs.org/isps/guide/filtering/

[28] Austin Murdock, Frank Li, Paul Bramsen, Zakir Durumeric, and Vern Paxson. 2017. Target Generation for Internet-Wide IPv6 Scanning. In *Proc. ACM Int. Measurement Conference (IMC)* (London, United Kingdom).

[29] University of Oregon. 2023. *University of Oregon Route Views Project.* http://www.routeviews.org/routeviews/

[30] Ramakrishna Padmanabhan, Zhihao Li, Dave Levin, and Neil Spring. 2015. UAv6: Alias Resolution in IPv6 Using Unused Addresses. In *Proc. Passive and Active Measurement (PAM)*.

[31] Craig Partridge and Mark Allman. 2016. Addressing Ethical Considerations in Network Measurement Papers. *Commun. ACM* 59, 10 (Oct. 2016).

[32] The ZMap Project. 2023. *ZGrab 2.0.* Retrieved 2023-03-24 from https://github.com/zmap/zgrab2

[33] Patrick Sattler, Johannes Zirngibl, Mattijs Jonker, Oliver Gasser, Georg Carle, and Ralph Holz. 2023. *Data and Analysis at TUM University Library.* https://mediatum.ub.tum.de/1723389 doi:10.14459/2023mp1723389.

[34] Patrick Sattler, Johannes Zirngibl, Mattijs Jonker, Oliver Gasser, Georg Carle, and Ralph Holz. 2023. *HRP Website with data.* Retrieved 2023-10-05 from https://hrp-stats.github.io/

[35] Khwaja Zubair Sediqi, Lars Prehn, and Oliver Gasser. 2022. Hyper-Specific Prefixes: Gotta Enjoy the Little Things in Interdomain Routing. *ACM SIGCOMM Computer Communication Review* 52 (June 2022). Issue 2. https://doi.org/10.1145/3544912.3544916

[36] Shadowserver. 2023. *Shadowserver - Lighting the way to a more secure Internet.* Retrieved 2023-03-24 from https://www.shadowserver.org/

[37] Shodan. 2023. *Shodan Dashboard.* Retrieved 2023-03-24 from https://www.shodan.io/dashboard

[38] Rapid7 Project Sonar. 2023. *Open Data.* Retrieved 2023-03-24 from https://opendata.rapid7.com/

[39] Roland van Rijswijk-Deij, Mattijs Jonker, Anna Sperotto, and Aiko Pras. 2016. A high-performance, scalable infrastructure for large-scale active DNS measurements. *IEEE journal on selected areas in communications* 34, 6 (2016), 1877–1888.

[40] Gerry Wan, Liz Izhikevich, David Adrian, Katsunari Yoshioka, Ralph Holz, Christian Rossow, and Zakir Durumeric. 2020. On the Origin of Scanning: The Impact of Location on Internet-Wide Scans. In *Proceedings of the ACM Internet Measurement Conference* (Virtual Event, USA) *(IMC '20)*. Association for Computing Machinery, New York, NY, USA, 662–679. https://doi.org/10.1145/3419394.3424214

[41] Johannes Zirngibl, Steffen Deusch, Patrick Sattler, Juliane Aulbach, Georg Carle, and Mattijs Jonker. 2022. Domain Parking: Largely Present, Rarely Considered!. In *Proc. Network Traffic Measurement and Analysis Conference (TMA) 2022*.

[42] Johannes Zirngibl, Patrick Sattler, and Georg Carle. 2023. A First Look at SVCB and HTTPS DNS Resource Records in the Wild. In *2023 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. 470–474.

[43] Johannes Zirngibl, Lion Steger, Patrick Sattler, Oliver Gasser, and Georg Carle. 2022. Rusty Clusters? Dusting an IPv6 Research Foundation. In *Proc. ACM Int. Measurement Conference (IMC)* (Nice, France).

(a) TCP/8443

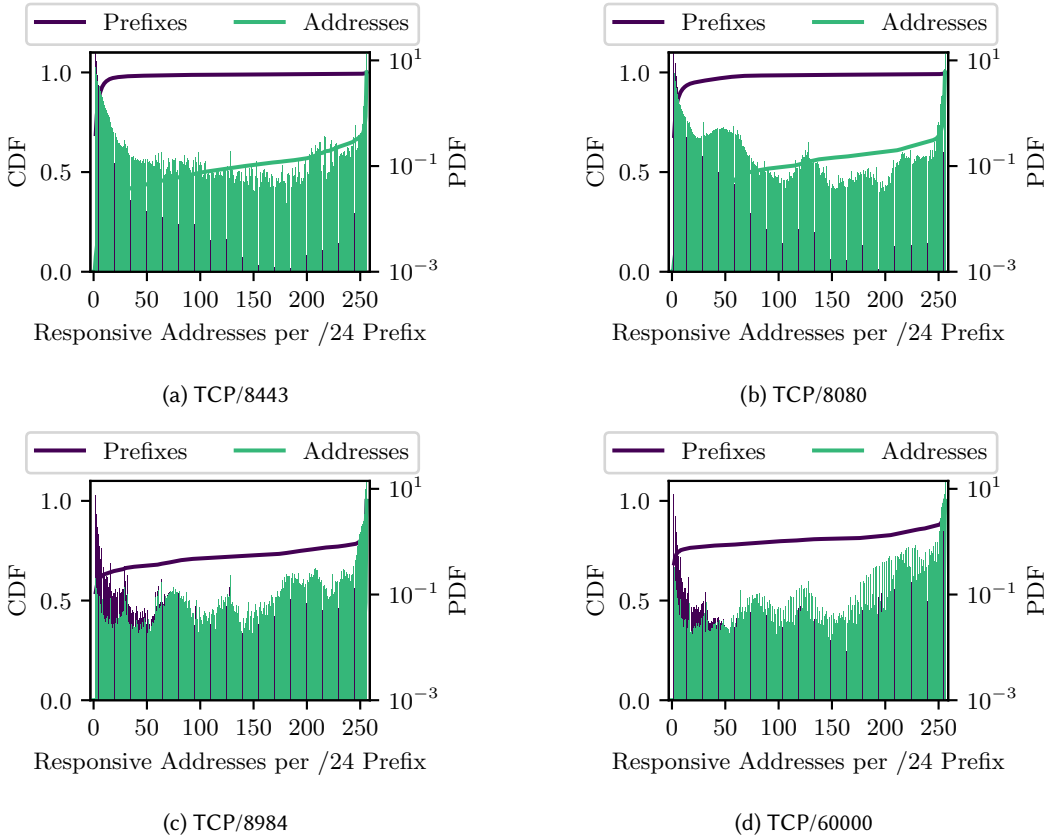(b) TCP/8080

(c) TCP/8984

(d) TCP/60000

Fig. 8. Respective and cumulative probability distribution of responsive addresses inside prefixes. The address data represents the impact of these prefixes on the scan results. Note the logarithmic axis for the right Y-axis.

A APPENDIX

Figure 8 shows the distribution of prefix responsiveness for the two most common HTTP/S alternative ports, TCP/60000, and TCP/8984. They show similar properties to Figure 1 and also confirm the joint evaluation in Figure 2. TCP/60000 and TCP/8984 exhibit an even more pronounced bias towards HRPs.