Exploring the Discovery Process of Fresh IPv6 Prefixes: An Analysis of Scanning Behavior in Darknet and Honeynet

Liang Zhao[1][0009−0002−9378−1864], Satoru Kobayashi[2][0000−0003−1017−0938], and Kensuke Fukuda[3][0000−0001−8372−2807]

[1] Sokendai, Tokyo, Japan
zhaoliang@nii.ac.jp
[2] Okayama University, Okayama, Japan
sat@okayama-u.ac.jp
[3] NII / Sokendai, Tokyo, Japan
kensuke@nii.ac.jp

Abstract. Internet-wide scanners can efficiently scan the expansive IPv6 network by targeting the active prefixes and responsive addresses on the hitlists. However, it is not clear enough how scanners discover fresh prefixes, which include newly assigned or deployed prefixes, as well as previously unused ones. This paper studies the whole discovery process of fresh prefixes by scanners. We implement four DNS-based address-exposing methods, analyze the arrival sequence of scans from distinct ASes, and examine the temporal and spatial scan patterns, with darknet and honeynet. Over six months, our custom-made darknet and probabilistic responsive honeynet collected 33M packets (1.8M sessions) of scans from 116 distinct ASes and 18.8K unique source IP addresses. We investigate the whole process of fresh prefix discovery, including address-exposing, initial probing, hitlist registration, and large-scale scan campaigns. Furthermore, we analyze the difference in scanning behavior by ASes, and categorize the scanners into three types, honeynet-exclusive, honeynet-predominant and balanced, based on the respective ratio of scans to darknet and honeynet. Besides, we analyze the intentions of scanners, such as network reconnaissance or scanning responsive targets, and the methods they used to obtain potential targets, such as by sending DNS queries or using public hitlist. These findings bring insights into the process of fresh prefixes attracting scanners and highlight the vital role of responsive honeynet in analyzing scanner behaviors.

Keywords: Network Security · IPv6 scan · Honeynet · Darknet.

1 Introduction

The Internet-wide scan is an important tool for benign and malicious activities to collect host and service information. With the recent adoption of IPv6 networks, several studies [10, 22, 28] report that IPv6 scans have been growing. However,

unlike the IPv4 network, scanning in the IPv6 network is much more difficult, because scanning the whole IPv6 space is impossible due to its vastness. Thus, the scanning activities in the IPv6 network are still not as prevalent as those in the IPv4 network [2, 6, 20, 21].

The hitlist [7, 11, 12, 15, 23, 30], containing active IPv6 prefixes and responsive addresses, enables scanners to perform scans efficiently across the expansive IPv6 network. A publicly available IPv6 hitlist is maintained by the Technical University of Munich (TUM) [15], though some scanners might rely on their lists. Scanners often target active IPv6 prefixes and responsive addresses on the hitlist. Thus, if a prefix or host is on the hitlist, it is more likely to be scanned. Also, target address generation [5, 9, 18, 24, 26, 29] have been well studied. However, even with the hitlist and target generation methods, it is hard for scanners to discover fresh IPv6 prefixes, which include newly deployed or assigned prefixes, or previously unused ones.

We characterize the whole process of fresh IPv6 prefixes being discovered by scanners. We introduce four DNS-based address-exposing methods to expose the target addresses to the Internet to attract the attention of scanners. We deploy multiple darknets and honeynets to measure the scanning activities associated with various address-exposing methods. A darknet is a network telescope that waits passively for the scans. A honeynet is a deceptive network that is purposely deployed to be scanned, attacked or compromised by the scanner and potential attacker [19, 25]. In this paper, we deploy probabilistic responsive honeynets, in which only a portion of addresses respond to scans, to prevent our honeynets from being recognized as aliased networks. More specifically, we want to answer the following questions. **RQ1:** How do scanners discover fresh IPv6 prefixes and what is the whole process of it? **RQ2:** How does the scanning behavior differ between darknet and honeynet? **RQ3:** What are the possible intentions of the scanners and how did they obtain the potential targets?

The contributions of our paper are as follows:

1. We design and develop an IPv6 scan detection system with darknet and probabilistic responsive honeynet, and deploy four DNS-based address-exposing methods to validate their effectiveness. We collected 33M packets (1.8M sessions) of scans from 116 distinct Autonomous Systems (ASes) and 18.8K unique source IP addresses over six months in 2023.
2. As the addresses of IPv6 fresh prefixes are exposed, it takes around 1-2 weeks before the scan begins to increase after the exposed prefixes are added to the hitlist. We analyze the sequence of arrival of scans from ASes and temporal scan patterns. We clarify the discovery process of fresh IPv6 prefixes as (1) address exposure, (2) DNS queries, (3) initial probes, (4) hitlist registration, and (5) follow-up scans.
3. Despite using the same exposing methods, responsive addresses in the honeynet received a greater amount of scans from a wider variety of source IP addresses and ASes compared to unresponsive addresses in the darknet. We compare the scanning behavior in the darknet and honeynet and categorize the scanners into three types, honeynet-exclusive (hitlist-based), honeynet-

predominant, and balanced (DNS registered-based and/or random), based on the respective ratio of scans to darknet and honeynet.
4. We analyze the intentions of scanners and the methods they use to obtain potential targets. Scanners targeting unregistered addresses probe the target network space with certain strategies for network reconnaissance, while those only targeting the registered addresses scan responsive targets specifically for exploitation. Scanners targeting registered addresses in both the darknet and honeynet are likely to obtain the registered addresses by sending the DNS queries to the auth server, while those only targeting responsive addresses in the honeynet are likely to obtain these addresses from the hitlist.

2 Scan Detection Methods

2.1 System Overview

We design and develop a system to detect the IPv6 scanning activities. We deploy multiple darknets (i.e., blackhole waiting passively for the scans) and honeynets (i.e, deceptive network purposely being deployed to be scanned or attacked [19, 25]) to compare the scanning behaviors within them. The maintainer of the hitlist employs detection algorithms to detect the aliased prefixes, in which all the addresses are responsive. We use probabilistic responsive honeynets, where only a portion of addresses actively respond to scans, to prevent them from being recognized as aliased networks. Our honeynet responds to ICMPv6 Echo, TCP SYN, and UDP requests. The target network in the experiment is a dedicated /56 IPv6 prefix, which is a subnet of a previously announced /32 prefix. Although the /56 prefix was not entirely newly assigned, it was not explicitly announced by BGP and was not used for other purposes. It is expected to experience a minimal level of internet background radiation before the experiment, similar to the unexposed darknet and honeynet (detailed in Tab. 1). We divide the /56 prefix into non-adjacent /64s and allocate them as darknets and honeynets, implementing different address-exposing methods in each, as shown in §2.2.

2.2 Address-Exposing Methods

We design and implement four DNS-based methods to expose addresses to the Internet. The DNS-based methods are proven to be effective in attracting scanners [28]. We conduct experiments to verify the effectiveness of different address-exposing methods and scenarios, comparing scans in darknets and honeynets. For each scenario, we allocate two /64s for the darknet and honeynet. We utilize the RFC-compliant DNS nameserver Name Server Daemon (NSD) [17] as our authoritative server.

IPv4 Reverse: Some IPv6 Scanners check the PTR records of IPv4 addresses and then use forward lookups to collect the associated IPv6 addresses for targets. We use domains associated with both IPv4 and IPv6 addresses to expose our IPv6 addresses. First, we register IPv4 PTR records to the auth

Table 1. Configurations and detected scans in each experiment

ID	Exposing Method (Scenario)	#Session	#SIPs	#ASes
dark-unex	No exposing	102	5	4
honey-unex	No exposing	5	1	1
dark-v4rev1	IPv4 reverse (1 random addr.)	23.1k	1.7k	13
honey-v4rev1	IPv4 reverse (1 random addr.)	330.9k	13.9k	37
dark-v4rev5	IPv4 reverse (5 random addr.)	109.4k	2.0k	13
honey-v4rev5	IPv4 reverse (5 random addr.)	1158.0k	16.0k	121
dark-v6enum	IPv6 enumeration (100 random addr.)	91	2	2
honey-v6enum	IPv6 enumeration (100 random addr.)	6	2	2
dark-v6spe	Special IPv6 (90 wordy & 18 port-embed)	8	1	1
honey-v6spe	Special IPv6 (90 wordy & 18 port-embed)	5	1	1
dark-pop	Popular service name (100 random addr.)	111	4	3
honey-pop	Popular service name (100 random addr.)	6	2	2

server to return the domain names to IPv4 reverse lookup queries. Then, we register AAAA records to the auth server to return the IPv6 addresses we want to expose to IPv6 forward lookup queries.

We set up two scenarios to verify the effect of the numbers of registered IPv6 addresses: (1) register one randomly chosen IPv6 address for each target subnet. (2) register five randomly chosen IPv6 addresses for each target subnet.

IPv6 Enumeration: Scanners can collect IPv6 addresses by exploiting the "denial of existence" semantics of the DNS (NXDOMAIN) [7]. This method involves iterating over the ip6.arpa zone, and performing a request for each possible child node, ignoring any subtree that returns a NXDOMAIN response. We register the IPv6 PTR records of 100 randomly chosen addresses from the target subnet to the IPv6 reverse lookup zone file so that these IPv6 addresses can be collected by the method in Ref. [7].

Special IPv6 Address: Some special addresses are often set manually and are more likely to be scanned according to RFC7707 [13]. The special addresses include those embedded with the service port of TCP/UDP services (e.g., 2001:db8::80), and addresses embedded with words (e.g., 2001:db8::cafe). We register 118 of these special addresses to the IPv6 reverse lookup zone file.

Popular Service Name: Domain name reconnaissance tools, such as fierce [8], use name dictionaries to scan the domain names of a service. We collect the 100 most popular words from the dictionaries of various domain reconnaissance tools and create domain names with these words, with our domain (in TLD ".com"). We register the IPv6 addresses to these domains so that these addresses are more likely to be scanned by scanners using domain name reconnaissance tools. We also made these domain names publicly available on the Internet so that they can be crawled or accessed by users, thereby increasing the likelihood that they will be discovered by scanners.

The experiment setup is shown in Table 1. As a control experiment, we set up two subnets with no address exposure, dark-unex and honey-unex.

2.3 Acquiring the Hitlist

We analyze the current and historical TUM hitlist [15] to determine whether our subnet and addresses have been included in the hitlist and when they were collected. The hitlist contains three separate lists: aliased prefixes, non-aliased prefixes and responsive addresses. The aliased prefixes list contains prefixes where all addresses are responsive, while non-aliased prefixes are those that are not aliased. The responsive addresses list contains the addresses that respond to ICMPv6, TCP or UDP requests. We discuss how our target network is included in the hitlist in detail in §3.

3 Validation of Methods

3.1 Dataset Overview

We deployed the scan detection system and started the experiment in Apr 2023. Our analysis is based on scan sessions. A scan session for TCP and UDP is defined as a 5-tuple bidirectional flow with a timeout, or maximum packet inter-arrival time, of 3,600 seconds. Similarly, a scan session of ICMPv6 Echo is a bidirectional flow that shares the same source/destination address, and request ID, with a timeout of 3,600 seconds. We set the timeout to 1 hour regarding Ref. [22]. Over six months, we collected 33M packets (1.8M sessions) of scans from 116 distinct ASes and 18.8K unique source IP addresses.

3.2 Effectiveness of Address-Exposing Methods

Table 1 shows the number of sessions received in each experiment, along with the number of unique source IPs and ASes. We show the information of a sample of ASes of IPv6 scanners in appendix A.

Comparison in the Four Address-Exposing Methods: The IPv4 reverse method attracted more than 99.99% of the scans, while other methods did not receive any associated DNS queries. Other methods (dark/honey-v6enum, v6spe, pop) only received minimal scans targeting the unregistered addresses, similar to the subnets with no exposure. This result suggests that the NXDOMAIN technique in the IPv6 enumeration method, brutal force dictionary attacks on specific IPv6 addresses or DNS reconnaissance are not yet fully exploited by the IPv6 scanners on a large scale. These methods might not be as straightforward, low-risk or cost-effective in terms of time and computing resources compared to the IPv4 reverse method to discover IPv6 addresses and prefixes. We will discuss the limitations of our address-exposing methods in §6. Dark-unex, v6enum, and pop received more scans than the others probably because their last nibble of the /64 prefix is 0 (e.g. 2001:db8:1:abc0::/64), which is easier to be selected for a scan. Interestingly, we identified two ASes (CERNET2(AS23910) and Alibaba Advertising(AS37963)) that were engaged in large-scale network reconnaissance spread in our target block (see also §4).

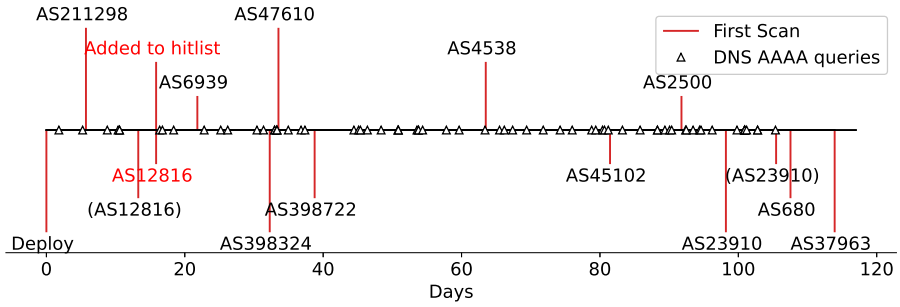

Fig. 1. The arrival sequence of the initial scan from each AS in the Darknet. An AS number enclosed in parentheses (e.g. (AS12816)) indicates the initial scan targeting an unregistered address from this AS. An AS number without parentheses indicates the initial scan targeting a registered address from this AS. See also Table 3 for AS names.

Scans Attracted by IPv4 Reverse Method: We compare the two scenarios in the IPv4 reverse method: one random IPv6 address mapping from one IPv4 address, and five random IPv6 addresses from one IPv4 address. In general, the honeynet attracted more scans than the darknet, similar to the trend in IPv4 Internet. The source IPs and ASes in the darknet of the two scenarios, dark-v4rev1 and dark-v4rev5, are almost identical, while the sessions of scan in dark-v4rev5 are 4.7 times greater than that in dark-v4rev1. The darknet with five registered addresses attracts more scans than the darknet with a single registered address, but these scans seem to originate from the same ASes and source IPs. On the other hand, the sessions in the honeynet of the two scenarios, honey-v4rev1 and honey-v4rev5, are almost identical, while the number of ASes of honey-v4rev5 is three times greater than that in honey-v4rev1 and the source IPs of honey-v4rev5 are 1.2 times greater than that in honey-v4rev1. The difference in sessions between honeynets of the two scenarios is not as substantial as observed in darknet. In summary, registering multiple IPv6 addresses for one IPv4 address is useful in amplifying sensors' resolution in the honeynet/darknet.

We will focus on the scans attracted by the IPv4 reverse method.

3.3 Arrival Sequence and Hitlist Confirmation

We investigate the arrival sequence of the ASes in the darknet and the honeynet, as shown in Figure 1 and 2, with DNS AAAA queries shown for reference.

DNS Queries: Regarding the DNS AAAA queries associated with the IPv4 reverse method, we confirmed the arrival of queries for the registered addresses in dark/honey-v4rev1/v4rev5 right after exposing them. However, we were unable to establish a correlation between the DNS queries with the initial scans from each AS, since the scanner may use cloud-hosting services for name resolution.

Hitlist Registration: As we noted in §2.3, the hitlist contains three separate lists, aliased prefixes, non-aliased prefixes and responsive addresses. We

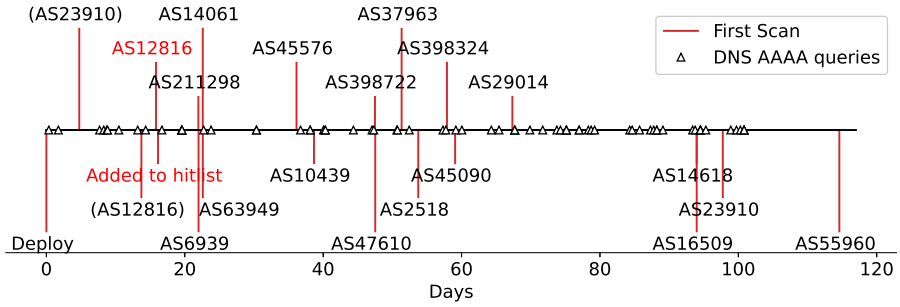

Fig. 2. The arrival sequence of the initial scan from each AS in the Honeynet.

confirmed that the target networks associated with IPv4 reverse (dark/honey-v4rev1, v4rev5) are added to the non-aliased *prefixes* lists and the registered honeynet addresses (1 in honey-v4rev1 and 5 in honey-v4rev5) are added to the responsive *address* hitlist about 17 days after deployment. Honeynet addresses responding to scans from Leibniz-Rechenzentrum(AS12816) are added to the responsive addresses list.

Arrival Sequence of ASes: Initially, the scanners probe the target network. Once prefixes are added to the hitlist, more scans originate from a diverse set of ASes. The ASes fall into two categories based on the target of their initial scan. Firstly, the initial scan targets an unregistered address, as seen with CERNET2 in the honeynet. This suggests that the scanner is conducting network reconnaissance. The registered address was scanned by this AS about 90 days later. Secondly, the initial scan targets the registered addresses, which are supposed to be discovered via hitlist or DNS queries (i.e., an intentional scanning activity). As the hitlist does not include the registered darknet addresses, the scanners targeting these addresses are likely to discover them via DNS queries independently or retrieve them from third parties.

In short, we confirmed the discovery process of fresh IPv6 prefixes: address exposure, DNS queries, initial probes, hitlist registration, and follow-up scans.

4 IPv6 Scan Characteristics

We analyze the scans received over time, and their sources, protocols and ports.

4.1 Scans over Time

Overall Trend: Figure 3 shows the hourly scan sessions over time in each subnet of the IPv4 reverse method since Apr 2023. Some periods of data are missing from Jun to Sep due to server shutdowns. In May and Jun, the daily scans in both the darknet and honeynet are regular; They are most active from 12am - 4pm UTC (peak 2 pm), though they significantly decreased from 4pm - 12am UTC. There

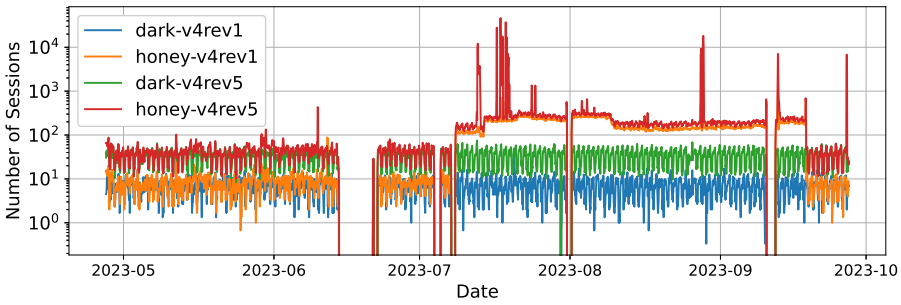

Fig. 3. Number of sessions received in each subnet using IPv4 reverse method

is no significant difference between weekdays and weekends. In both darknet and honeynet, a larger number of scans are observed for a larger number of registered addresses. In Jul, we observed a strong uptick of scans in the honeynets, where a large-scale scan campaign has begun spanning over two months. The average hourly scans in honey-v4rev1 and honey-v4rev5 have increased to a few hundred since Jul. There are also spikes of scans in honey-v4rev5 from Jul to Sep. In the middle of Sep, the large-scale scan campaign ended and the daily scan in the honeynet dropped to the same level as May and Jun. However, the spikes in honey-v4rev5 can still be observed.

Large-Scale Scans: As in Ref. [22], we define a large-scale scan as a source targeting at least 100 destination IPv6 addresses with a timeout of 1 hour. According to this definition, we have also observed 28 large-scale scan events from five different sources in our dataset. In a single scan, more than 20,000 unique destination addresses are targeted at most. Most of the destinations of these large-scale scans are in the honeynet, which indicates that the reactive responses might be a triggering factor of large-scale scans.

4.2 Scan Targets

Next, we analyze the scans by their targets: the registered addresses or unregistered addresses. Scans targeting registered addresses account for 97.67% of scans in the darknet and 52.03% of scans in the honeynet. Most of the remaining scans in the honeynet target unregistered addresses near the registered address, accounting for 47.58% of the total scans in the honeynet. However, the remaining scans in the darknet are targeting random unregistered addresses across the /64 subnet, rather than those near the registered address. Table 1 shows the percentage of scans targeting the registered addresses for each AS. Scanners targeting only registered addresses are scanning these addresses specifically for exploitation, while those targeting both registered and unregistered addresses are likely conducting network reconnaissance to discover previously unknown active hosts. Scans targeting the lower byte IID in the /64 subnet (e.g., 2001:db8::2) only account for 0.03% of the total scans.

4.3 Scan Sources

Table 2. Top source ASes targeting Darknet and Honeynet, sorted by #Session.

Darknet					Honeynet				
ASN	**#Session**	**Regist**	**/128s**	**/64s**	**ASN**	**#Session**	**Regist**	**/128s**	**/64s**
6939	66.9K	100.0%	1.7k	3	23910	685.5k	0.1%	5	2
211298	61.1K	100.0%	231	7	16509	469.3k	100.0%	9.3k	172
12816	3.6k	16.9%	9	1	14618	152.8k	100.0%	2.3k	17
3910	268	62.7%	3	1	6939	67.6k	100.0%	1.7k	4
37963	207	99.0%	2	1	211298	61.4k	100.0%	236	7
398722	199	100.0%	13	1	37963	21.1k	1.4%	14	13
398324	151	100.0%	14	1	10439	13.3k	100.0%	11	1
47610	99	100.0%	1	1	10439	13.3k	100.0%	11	1
2500	35	100.0%	6	2	2637	2.1k	6.4%	1	1
680	24	100.0%	1	1	14061	1.7k	99.9%	579	12

Next, we analyze the intentions of scanners and the methods they use to obtain the potential targets by their sources. Table 2 lists the top source ASes targeting the darknet and honeynet. The top source ASes include R&E networks (CERNET2), ISPs (Hurricane Electric), cloud hosting service (AMAZON-02(AS16509)) and business firms (Alibaba Advertising). The distribution of scan sessions from each AS is long-tailed. The majority of scans originate from a few ASes in both the darknet and honeynet. In the darknet, the scans of the top three ASes account for a combined total of 99% of all scans. In the honeynet, the scans from the top three ASes contribute to a combined total of 88% of all scans. We can only identify two scanners from Censys [3] in an open access acknowledged scanners repository [4].

Scanner Types: There are 13 ASes that simultaneously scan both the darknet and honeynet, 94 ASes exclusively scan the honeynet, while there is no AS that exclusively scans the darknet. Figure 4 shows the scan sessions towards darknet and honeynet for ASes. The ASes can be categorized into three types: honeynet-predominant, honeynet-exclusive and balanced.

Honeynet-Predominant: For ASes above the diagonal such as CERNET2 and Alibaba Advertising, the number of scans in the honeynet is significantly higher than that in the darknet. These ASes probe both the darknet and honeynet initially and then conduct large-scale scans toward the honeynet. As we mentioned in §3.2, they contribute to the majority of scans to unregistered addresses all across our /56 target network, regardless of the address-exposing method or scenario. Moreover, these scanners mainly target unregistered addresses near the registered address. This indicates that their purpose is to conduct network reconnaissance to find potential targets and then launch large-scale scans targeting these active hosts.

Honeynet-Exclusive: The ASes on the left side of the figure exclusively scan the honeynet. Most of these ASes target the registered addresses exclusively, suggesting they may have learned the registered responsive addresses via the

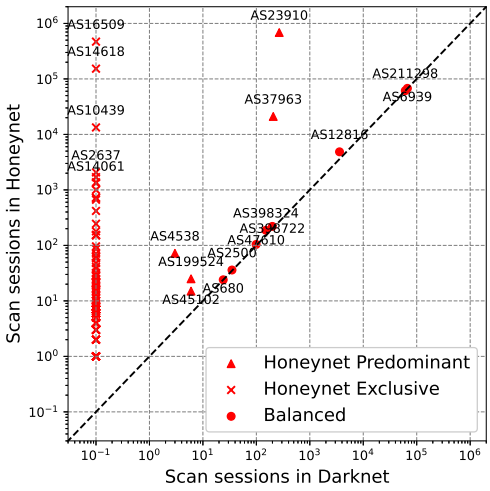

Fig. 4. Scatter plot of Darknet and Honeynet for each AS (10^{-1} indicates no sessions).

hitlist[4] and were scanning these addresses intentionally. During large-scale scan campaigns, AMAZON-02 and AMAZON-AES(AS14618) conduct scan sessions that last for more than 30 hours, generating over 10,000 packets. However, there are still exceptions like GEORGIA-TECH(AS2637), which scans unregistered addresses, suggesting searching for scanning targets.

Balanced: For the ASes along the diagonal, the number of scans is nearly identical for both the darknet and honeynet. The scans are usually stable over time with a fixed periodic cycle. This suggests that the purpose of these ASes scan the network thoroughly without any specific preferences. They mainly target the registered address (e.g., Constantine Cybersecurity(AS211298) and Hurricane Electric), or random unregistered addresses in the /64 subnet (e.g., Leibniz-Rechenzentrum).

In conclusion, honeynet-predominant scanners target the responsive addresses after a probing phase and also search for potential targets near the registered addresses. Balanced scanners probe the entire network space and target the registered addresses or random unregistered addresses. Honeynet-predominant scanners and balanced scanners target registered addresses in both the honeynet and darknet. Therefore, they likely obtain the targets by sending the DNS queries to the authoritative DNS server, since the hitlist only contains responsive addresses. Honeynet-exclusive scanners primarily target the registered addresses in the Honeynet, and they likely obtain these addresses from the hitlist.

Scans over Time by AS in Honeynet: Figure 5 shows the scan over time by AS in honey-v4rev5. Before the uptick in scans in July, the majority of the scans were from security firms and research institutions. These ASes equally scan the darknet and honeynet, showing a periodic fluctuation pattern over time.

[4] Recall that the TUM hitlist does not provide registered darknet addresses.

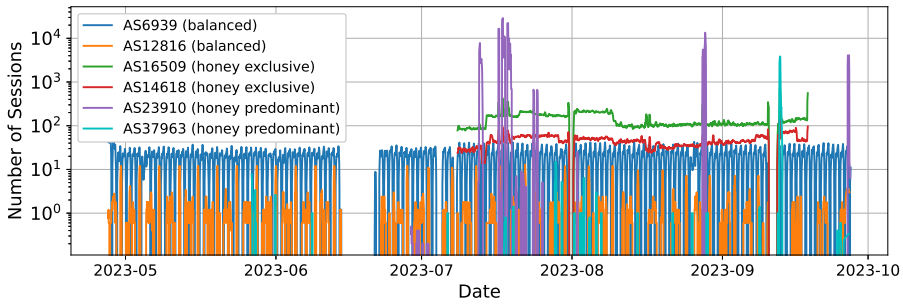

Fig. 5. Number of sessions from each source AS received in honey-v4rev5

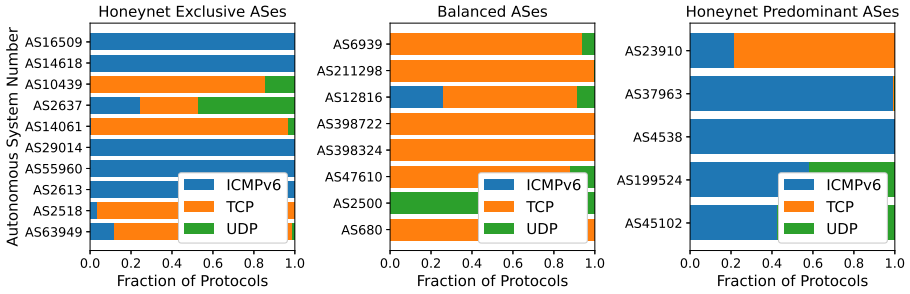

Fig. 6. Fraction of protocols of darknet and honeynet

Scans from Hurricane Electric show a consistent daily pattern, with scanning activities only occurring during the hours of 0am - 3pm UTC, and no scans for the rest of the day. Scans from Leibniz-Rechenzentrum show a notable large surge of scans every four days, with some smaller surges in between. Scans from AMAZON-02 and AMAZON-AES show identical patterns, indicating that they are likely to be carried out by the same entity, while AMAZON-02 has a 4 times/128s and 10 times/64s source than AMAZON-AES. In contrast, scans from CERNET2 show significant surges, which mainly target addresses that are not registered on the DNS server. The scans from other ASes only target registered addresses.

4.4 Scan Protocols and Target Ports

Protocols: We analyze the fraction of protocols of scans in each AS and compare the honeynet-exclusive, honeynet-predominant and balanced ASes. Figure 6 shows the fraction of protocols of each type of ASes. In honeynet-exclusive ASes, many scanners use only ICMPv6. On the other hand, balanced ASes use primarily TCP. Most of the honeynet-predominant ASes use primarily ICMPv6, along with TCP or UDP. Interestingly, Leibniz-Rechenzentrum and GEORGIA-TECH, two research institutions, use a combination of TCP, UDP and ICMPv6 to get a whole landscape of the IPv6 network for research proposes.

Ports: Table 4 and 5 in appendix B show the top destination ports and number of unique ports of the darknet and the honeynet. Some scanners primarily focused on well-known and commonly used ports (i.e., 22, 53, 80, 443 and 8080), such as Leibniz-Rechenzentrum and Hurricane Electric(AS6939). We also observed penetration test behaviors, which scan across a wide range of ports, from some scanners. For example, Constantine Cybersecurity is scanning over 4,000 unique target ports in both darknet and honeynet. Interestingly, both Alibaba Advertising and CERNET2 are targeting the TCP/7547 port, a commonly used target port for the Mirai botnet and its variants [1]. These two ASes are using the same source port, specifically port TCP/37492, for their scanning activities in the darknet, which suggests that these two ASes could either be the same entity or they might use the same scanning software.

5 Related Work

Scan Detection: Scan detection is diligently carried out in the IPv4 network [2, 6, 20, 21]. Darknet and honeynet [16, 19, 25] are frequently employed to collect information about scanning activities. Recent works [10, 14, 22, 27, 28] indicate the increasing prevalence of scanners covering the IPv6 space. Richter et al. [22] investigated the large-scale IPv6 scanning activities on the Internet based on firewall logs captured at a major CDN. They studied the dominant characteristics of scans, including scanner origins and targeted services. Their work focuses more on already existing scan activities in the broad internet space, while we focus on the growth of scan activities in a fresh prefix that is dedicated to our experiments and not used for other purposes. Furthermore, the detected scans in our work are attracted to the target network by address exposing methods, while the scans in Ref. [22] are general scans observed in a CDN. The scans we've detected in our work are supposed to be more specific and exhibit greater similarity. This allows us to analyze their common characteristics and purposes.

Scan Attraction: Tanveer et al. [28] propose to attract potential IPv6 scans by direct/indirect scan attraction methods. Among their indirect (passive) contact approaches, DNS-based (DNS zone) and NTP-based methods (NTP pool/public) are effective in attracting the scans. They studied the relationship between the attraction methods and the scanner attention they evoke. We took a step further and did a more thorough and focused analysis of the scans attracted by the same exposing method. We focused on the DNS-based approach and differentiated the scanner behavior based on their source ASes, targeting (exposed addresses or random addresses), preference on darknet or honeynet, and traffic patterns over time. We show that even the scans attracted by the same exposing method could still vary significantly in their behavior.

Address Discovery: IPv6 hitlist-based scans [7,11,12,15,23,30] have emerged, enabling scanners to perform scans across the vast IPv6 network space. Target generation methods [5, 9, 18, 24, 26, 29], which generate target addresses based on known addresses, have been proposed to find IPv6 addresses to scan.

6 Limitations

There are still some limitations in this work. First, we mainly focused on DNS-based address exposing methods and the dataset we collected is primarily attracted through the IPv4 reverse method. This dataset allows us to perform a thorough and focused analysis of the scans attracted by the same method. However, address-exposing methods using other services [28], such as NTP-based methods (NTPpool, NTP public), can also attract scans effectively. We plan to implement other methods to retrieve a more complete dataset in future work. Second, we have not observed any DNS queries associated with IPv6 enumeration and special IPv6 methods, possibly due to an insufficient measurement period to fully expose the addresses. Moreover, IPv6 enumeration requires that authoritative DNS servers for parent zones are also RFC-compliant, which is not always guaranteed. Regarding the popular service name method, the URLs we made public are not drawing significant Internet traffic, and are not yet indexed by Google or other search engines. Moreover, we added the name of the experiment as a subdomain to the URLs, which could also make it harder for DNS reconnaissance tools to discover the full URLs. Third, distinguishing whether scans originate from the same user or not can be challenging, especially for multi-tenant cloud networks. However, we find some evidence showing that the scans from some of the ASes are from the same user. For example, both Alibaba Advertising and CERNET2 only use the same source port TCP/37492. AMAZON-02 and AMAZON-AES are sharing a similar scan traffic pattern. Lastly, we confirmed that our prefixes and addresses are added to the TUM hitlist. However, it is uncertain if the scanners are also using other hitlists. To address this issue, we consider intentionally not responding to TUM scans to avoid being included in their hitlist and analyze the effect of other hitlists in future work.

7 Conclusion

This paper studies the whole process of fresh prefixes being discovered by scanners, including address-exposing, initial probing, hitlist registration, and large-scale scan campaigns. We analyze the difference in scanning behavior by ASes and categorize the scanners into three types, honeynet-exclusive, honeynet predominant and balanced, based on the respective ratio of scans to darknet and honeynet. Besides, we analyzed the intentions of scanners, such as network reconnaissance or scanning responsive targets, and the methods they used to obtain potential targets, such as by sending DNS queries or using the hitlist. These findings bring insights into the process of fresh prefixes attracting scanners and highlight the vital role of responsive honeynet in analyzing scanner behaviors.

Acknowledgments. We thank our shepherd, Pawel Foremski, and anonymous reviewers for their valuable feedback and suggestions, which greatly improved the quality of our manuscript. This work is financially supported by JSPS 21H03438.

A ASes Information

We list information about a sample of the ASes of IPv6 scanners we confirmed in Table 3.

Table 3. The Description, ASN Type, Country, and Scanner Type of ASes

ASN	Description	ASN Type	CC	Scanner Type
16509	AMAZON-02	Hosting	US	Honeynet-Exclusive
14618	AMAZON-AES	Hosting	US	Honeynet-Exclusive
55960	Beijing Guanghuan Xinwang Digital	ISP	CN	Honeynet-Exclusive
10439	CARINET	Hosting	US	Honeynet-Exclusive
398324	CENSYS-ARIN-01	Business	US	Balanced
398722	CENSYS-ARIN-03	Business	US	Balanced
4538	China R&E Network Center	Research	CN	Honeynet-Predominant
23910	China Next Gen. Internet CERNET2	Research	CN	Honeynet-Predominant
211298	Constantine Cybersecurity	Business	UK	Balanced
14061	DIGITALOCEAN-ASN	Hosting	US	Honeynet-Exclusive
2637	GEORGIA-TECH	Research	US	Honeynet-Exclusive
6939	HURRICANE	ISP	US	Balanced
37963	Hangzhou Alibaba Advertising	Business	CN	Honeynet-Predominant
12816	Leibniz-Rechenzentrum	Research	DE	Balanced
47610	RWTH Aachen University	Research	DE	Balanced
29014	ScaleUp Technologies	Hosting	DE	Honeynet-Exclusive
680	Verein zur Foerderung eines Deutschen	Research	DE	Balanced
2500	WIDE Project	Research	JP	Balanced

B Port Information

Here, we provide the details of destination ports for darknet (Table 4) and honeynet (Table 5).

References

1. Antonakakis, M., April, T., Bailey, M., Bernhard, M., Bursztein, E., Cochran, J., Durumeric, Z., Halderman, J.A., Invernizzi, L., Kallitsis, M., Kumar, D., Lever, C., Ma, Z., Mason, J., Menscher, D., Seaman, C., Sullivan, N., Thomas, K., Zhou, Y.: Understanding the mirai botnet. In: Proc. USENIX Security'17. pp. 1093–1110. Vancouver, BC (Aug 2017)
2. Benson, K., Dainotti, A., claffy, k., Snoeren, A.C., Kallitsis, M.: Leveraging internet background radiation for opportunistic network analysis. In: Proc. ACM IMC'15. p. 423–436. Tokyo, Japan (Oct 2015)
3. Censys: Opt out of data collection, https://support.censys.io/hc/en-us/articles/360043177092-Opt-Out-of-Scanning, accessed in January 2024

Table 4. Primary ports of scans towards the Darknet. (Percentage) is the portion of scans to a certain port among all scans. Unique is the number of unique dest ports.

ASN	Top 3 Destination Ports						Uniq
6939	8081	(2.5%)	8080	(2.5%)	8001	(2.5%)	82
211298	8443	(0.5%)	21	(0.5%)	135	(0.5%)	4232
12816	80	(85.6%)	443	(10.6%)	161	(1.9%)	4
23910	443	(25.0%)	22	(25.0%)	7547	(25.0%)	4
37963	443	(25.0%)	22	(25.0%)	7547	(25.0%)	4
398722	443	(65.3%)	80	(34.7%)	–	–	2
398324	443	(66.2%)	80	(33.7%)	–	–	2
47610	5671	(30.3%)	8883	(18.2%)	55672	(18.2%)	7
2500	53	(100.0%)	–	–	–	–	1
680	443	(50.0%)	80	(50.0%)	–	–	2

Table 5. Primary ports of scans towards the Honeynet

ASN	Top 3 Destination Ports						Uniq
23910	443	(25.1%)	7547	(25.1%)	80	(25.1%)	26
6939	8080	(2.5%)	8888	(2.5%)	8001	(2.5%)	82
211298	443	(0.5%)	135	(0.5%)	22	(0.5%)	4171
37963	80	(24.3%)	443	(23.3%)	7547	(23.3%)	5
10439	2152	(2.6%)	2123	(2.1%)	53	(1.4%)	125
12816	80	(69.7%)	443	(22.5%)	161	(3.9%)	4
2637	53	(62.3%)	443	(20.9%)	80	(16.8%)	3
14061	443	(8.5%)	8443	(5.7%)	27017	(3.0%)	224
2518	443	(16.7%)	80	(14.6%)	8080	(13.7%)	7
63949	8443	(14.6%)	443	(11.1%)	5002	(2.4%)	165

4. Collins, M.P., Hussain, A., Schwab, S.: Identifying and differentiating acknowledged scanners in network traffic. In: 2023 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW). pp. 567–574 (2023). https://doi.org/10.1109/EuroSPW59978.2023.00069
5. Cui, T., Gou, G., Xiong, G.: 6gcvae: Gated convolutional variational autoencoder for ipv6 target generation. In: Proc. PAKDD'20. Singapore (May 2020)
6. Durumeric, Z., Bailey, M., Halderman, J.A.: An internet-wide view of internet-wide scanning. In: Proc. USENIX Security. p. 65–78. San Diego, CA (Aug 2014)
7. Fiebig, T., Borgolte, K., Hao, S., Kruegel, C., Vigna, G.: Something from nothing (there): Collecting global ipv6 datasets from dns. In: Proc. PAM'18. pp. 30–43 (Mar 2018)
8. Fierce: A DNS reconnaissance tool for locating non-contiguous IP space., https://github.com/mschwager/fierce, accessed in August 2023
9. Foremski, P., Plonka, D., Berger, A.: Entropy/ip: Uncovering structure in ipv6 addresses. In: Proc. ACM IMC'16. p. 167–181. Santa Monica, CA (Nov 2016)
10. Fukuda, K., Heidemann, J.: Who knocks at the ipv6 door? detecting ipv6 scanning. In: Proc. ACM IMC'18. p. 231–237. Boston, MA (Oct 2018)
11. Gasser, O., Scheitle, Q., Foremski, P., Lone, Q., Korczyński, M., Strowes, S.D., Hendriks, L., Carle, G.: Clusters in the expanse: Understanding and unbiasing ipv6 hitlists. In: Proc. ACM IMC'18. p. 364–378. Boston, MA (Oct 2018)

12. Gasser, O., Scheitle, Q., Gebhard, S., Carle, G.: Scanning the ipv6 internet: Towards a comprehensive hitlist. CoRR **abs/1607.05179** (2016), http://arxiv.org/abs/1607.05179
13. Gont, F., Chown, T.: Network Reconnaissance in IPv6 Networks. Tech. rep., Internet Engineering Task Force (Dec 2015), https://tools.ietf.org/html/rfc7707, RFC7707
14. Hiesgen, R., Nawrocki, M., King, A., Dainotti, A., Schmidt, T.C., Wählisch, M.: Spoki: Unveiling a new wave of scanners through a reactive network telescope. In: 31st USENIX Security Symposium (USENIX Security 22). pp. 431–448. USENIX Association, Boston, MA (Aug 2022), https://www.usenix.org/conference/usenixsecurity22/presentation/hiesgen
15. IPv6 Hitlist Service, https://ipv6hitlist.github.io/, accessed in August 2023
16. Javadpour, A., Ja'Fari, F., Taleb, T., Benzaid, C.: A mathematical model for analyzing honeynets and their cyber deception techniques. In: Proc. ICECCS (Jun 2023)
17. Labs, N.: The nlnet labs name server daemon (nsd) is an authoritative, rfc compliant dns nameserver., https://github.com/NLnetLabs/nsd, accessed in January 2024
18. Murdock, A., Li, F., Bramsen, P., Durumeric, Z., Paxson, V.: Target generation for internet-wide ipv6 scanning. In: Proc. ACM IMC'17. p. 242–253. London, UK (Nov 2017)
19. Nawrocki, M., Wählisch, M., Schmidt, T.C., Keil, C., Schönfelder, J.: A survey on honeypot software and data analysis (2016), http://arxiv.org/abs/1608.06249
20. Pang, R., Yegneswaran, V., Barford, P., Paxson, V., Peterson, L.: Characteristics of internet background radiation. In: Proc. ACM IMC'04. p. 27–40 (Oct 2004)
21. Richter, P., Berger, A.: Scanning the scanners: Sensing the internet from a massively distributed network telescope. In: Proc. ACM IMC'19. p. 144–157. Amsterdam, Netherlands (Oct 2019)
22. Richter, P., Gasser, O., Berger, A.: Illuminating large-scale ipv6 scanning in the internet. In: Proc. ACM IMC'22. p. 410–418. Nice, France (Oct 2022)
23. Rye, E., Levin, D.: Ipv6 hitlists at scale: Be careful what you wish for. In: Proceedings of ACM SIGCOMM'23. pp. 904–916 (Sep 2023)
24. Song, G., Yang, J., Wang, Z., He, L., Lin, J., Pan, L., Duan, C., Quan, X.: Det: Enabling efficient probing of ipv6 active addresses. IEEE/ACM Transactions on Networking **30**(4), 1629–1643 (Aug 2022). https://doi.org/10.1109/TNET.2022.3145040
25. Spitzner, L.: The honeynet project: trapping the hackers. IEEE Security & Privacy **1**(2), 15–23 (Mar 2003)
26. Steger, L., Kuang, L., Zirngibl, J., Carle, G., Gasser, O.: Target acquired? evaluating target generation algorithms for ipv6. In: Proc. TMA'23 (Jun 2023)
27. Strowes, S.D., Aben, E., Wilhelm, R., Obser, F., Stagni, R., Formoso, A.: Debogonising 2a10: : /12: Analysis of one week's visibility of a new /12. In: Proc. TMA'20 (Jun 2020)
28. Tanveer, H.B., Singh, R., Pearce, P., Nithyanand, R.: Glowing in the dark uncovering ipv6 address discovery and scanning strategies in the wild. In: Proc. USENIX Security'23. pp. 6221–6237 (Aug 2023)
29. Yang, T., Hou, B., Cai, Z., Wu, K., Zhou, T., Wang, C.: 6graph: A graph-theoretic approach to address pattern mining for internet-wide ipv6 scanning. Computer Networks **203**, 108666 (Feb 2022)

30. Zirngibl, J., Steger, L., Sattler, P., Gasser, O., Carle, G.: Rusty clusters? dusting an ipv6 research foundation. In: Proc. ACM IMC'22. p. 395–409. Nice, France (Oct 2022)