How Did You Find That Address? An In-Depth Investigation of Domain-Based IPv6 Scanners

SEBASTIAN KAPPES, Max Planck Institute for Informatics, Germany
LION STEGER, Technical University of Munich, Germany
VICTOR LE POCHAT, Independent, Belgium
OLIVER GASSER, IPinfo, United States
JOHANNES ZIRNGIBL, Max Planck Institute for Informatics, Germany
TIAGO HEINRICH, Max Planck Institute for Informatics, Germany

IPv6's vast address space makes scanning particularly challenging. IPv6 scanners must first discover in-use IPv6 addresses they can target and scan. One way of doing this is by collecting IPv6 addresses from *domain names*: if a DNS name resolves to an IPv6 address, that address is more likely to be used by an actual live host.

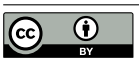
In this study, we investigate how IPv6 scanners rely on domain names for target reconnaissance. We set up a large IPv6 telescope to which we attract scanners by placing honey domain names on TLD zone files, CT logs, and the Tranco Top 1M. We introduce a technique to trace the packets that IPv6 scanners send to the telescope back to the initial DNS queries they sent to discover the IPv6 addresses, allowing us to better evaluate how large-scale IPv6 scanners collect and scan addresses. Over a period of six months, we collect 9.4 M IPv6 scan packets from 25 k addresses in 613 ASes. We find that domain-based IPv6 scanners differ in their address collection strategies, which also correlates with different scanner behavior: scanners that scan addresses hours after collecting them send mostly ICMP traffic, while scanners that resolve domains and scan addresses in one go are more focused on TCP/UDP—many of the latter are also not limited to IPv6 and will fall back to IPv4 if necessary. We uncover that scanners often combine (in particular, open) domain sources to find more targets, and that different domain sources attract different scanners. We also demonstrate that the IPv6 Hitlist is a strong mediating factor when attracting scanners through domain names, which we address by excluding our telescope's address space from the Hitlist. Furthermore, we use our fine-grained packet tracing to detect distributed scanning campaigns involving multiple ASes and to distinguish different scanners within large cloud ASes. Our findings show a diverse landscape of domain-based IPv6 scanning on the Internet: different scanners employ distinct scanning strategies and rely on unique domain sources, which greatly impacts how researchers should study them.

CCS Concepts: • **Security and privacy → Network security**; • **Networks → Network measurement**.

Additional Key Words and Phrases: IPv6, Internet scanning, network telescopes, darknets, domain names, Hitlist

ACM Reference Format:
Sebastian Kappes, Lion Steger, Victor Le Pochat, Oliver Gasser, Johannes Zirngibl, and Tiago Heinrich. 2026. How Did You Find That Address? An In-Depth Investigation of Domain-Based IPv6 Scanners. *Proc. ACM Netw.* 4, CoNEXT3, Article 35 (September 2026), 25 pages. https://doi.org/10.1145/3830390

Authors' Contact Information: Sebastian Kappes, Max Planck Institute for Informatics, Saarbrücken, Germany; Lion Steger, Technical University of Munich, Munich, Germany; Victor Le Pochat, Independent, Leuven, Belgium; Oliver Gasser, IPinfo, Seattle, United States; Johannes Zirngibl, Max Planck Institute for Informatics, Saarbrücken, Germany; Tiago Heinrich, Max Planck Institute for Informatics, Saarbrücken, Germany.



ACM 2834-5509/2026/9-ART35
https://doi.org/10.1145/3830390

1 Introduction

Scanning traffic on the Internet is ubiquitous. Internet hosts not only receive visits from ordinary users, but also from a myriad of bots and scanners [5, 9, 26, 30, 52]. Some are operated by benign entities such as security companies or researchers, others belong to malicious actors looking for vulnerable hosts [5, 9]. Understanding and studying *how* Internet scanners operate and *what intentions* they have constitutes an essential first step towards protecting against potential attacks.

In IPv4, conducting large-scale Internet scans is easy: a scanner that wants to do a port scan of every host on the IPv4 Internet can simply send probing packets to *all* IPv4 addresses and observe which hosts respond [10]. In IPv6, however, things are not straightforward. As the IPv6 address space is substantially larger than in IPv4, brute forcing all addresses is no longer feasible [15]. Instead, both benign and malicious actors trying to scan IPv6 must focus their scans on those parts of the address space that likely contain hosts—i.e., that show signs of *liveness* [2]. There are several possible indications that a portion of address space contains hosts, for example packets emanating from the network [37], Border Gateway Protocol (BGP) announcements of specific prefixes in the network [19, 38], or the presence of *domain names* [14]: if a domain name resolves to an IPv6 address, that address is likely in use, meaning that if the scanners send a port scan packet to the address, there is a greater chance that it reaches a responsive host.

But not only is *scanning* the IPv6 Internet harder: *studying* IPv6 scanners is also more difficult. In IPv4 security researchers often use network telescopes—sometimes also called darknets—to study IPv4 scanners [1, 5, 9, 12, 16, 28, 29, 31, 51]. Telescopes are large portions of unused address space that do not contain any live hosts. As one would usually not expect any traffic to such a network, the unsolicited probing packets of Internet-wide scanners will stand out, such that they can easily be collected and analyzed. Conversely, in IPv6, merely collecting scan traffic at a telescope is often not enough. Since there are no active hosts in a network telescope (i.e., since the address space is "dark"), there are no signs of liveness that encourage IPv6 scanners to target it. Without additional measures, IPv6 telescopes hence attract very little scan traffic compared to their IPv4 counterparts [7, 47]. Consequently, recent works have attempted to study IPv6 scanners by setting up network telescopes and, furthermore, signaling *liveness* of the address space to attract IPv6 scanners [11, 48, 49, 54]. One way to signal to scanners that the address space might contain active hosts is to have domain names point to the telescope. Initial works have shown that setting up a domain name leads to an increase of IPv6 scan traffic [11, 32, 48, 49, 54]. Nevertheless, apart from the fact that domain names can attract IPv6 scanners, not much is known about *what kind* of scanners this attracts, what their intentions may be, and how they operate.

Moreover, there are several pitfalls when analyzing IPv6 scan traffic that was attracted through domain scanners. Firstly, as one exposes IPv6 addresses through domain names to signal liveness of the network, it is unclear if the scan traffic attracted is IPv6-specific, or if scanners do not care whether their connection is established over IPv4 or IPv6. For example, by setting up domain names that resolve to IPv6 addresses in one's network, one will also attract web crawlers that are interested in the domains themselves, regardless of the underlying layer three protocol. The scanners found in IPv6 telescopes can therefore be either *IPv6-specific*, meaning they are specifically resolving domain names to discover IPv6 networks and hosts, or *layer-three-agnostic*, as in the case of the web crawlers that try to access the webpage under the domain. Also, mediating factors can give a false impression of what attracts IPv6 scanners. For example, if an address from a network telescope finds its way onto a popular IPv6 hitlist [14, 55] (a published list of IPv6 addresses confirmed to be live), one will see an increase in scanners targeting this address. These scanners, however, will have found the address through ways unrelated to the intended exposure methods.

Furthermore, since IPv6 address space is cheap, IPv6 scanners can distribute their scans, running them from multiple addresses in the same network. This makes it hard to tell how many distinct entities are scanning from within a specific network. Moreover, some organizations might run distributed IPv6 scans from vantage points in multiple networks, which can easily be misidentified as scans from different actors. Without further information, there is no way to tell if patterns in scan traffic are because of a single entity or multiple scanners.

The main idea of this paper is to design an IPv6 telescope that can overcome these obstacles and thereby substantially improve our understanding of domain-based IPv6 scanners. Specifically, we develop a novel *DNS-aware* IPv6 telescope: we set up a /36 IPv6 telescope and attract IPv6 scanners to it by exposing honey domain names on three different domain sources (Top Level Domain (TLD) zone files, Certificate Transparency (CT) logs, and top lists). However, when we attract scanners through domain names, we use a dedicated Domain Name System (DNS) setup that hands out a unique address from the telescope's address space for each DNS query it receives, thus allowing us to trace back scans in the telescope—possibly from different source addresses—to individual DNS queries. Additionally, we are the first to study domain-based IPv6 scanners in isolation *without* any of the mediating effects of popular public hitlists, which we show can have a substantial impact on IPv6 telescope traffic.

Through this study, we pose four research questions and provide the following contributions:

(i) **Can we distinguish domain-based IPv6 scanners based on their address collection strategies, and do scanners with different strategies behave differently?** Using our *DNS-aware* IPv6 telescope, we discover that scanners can be categorized based on whether they collect and scan addresses in one go or in distinct phases (Section 5.2). We show that one-go and dual-phase scanners are fundamentally different: they scan different protocols, originate from different Autonomous System (AS) types, and—in the case of one-go scanners—are often not restricted to IPv6 (Sections 5.4 and 5.5).

(ii) **How can we identify distributed scans from multiple sources that were conducted by the same entity?** Through our method, we are able to uncover related IPv6 scanners from both research institutions and unknown actors that were previously unidentified, showing that those can lead to overestimating the total number of scanners attracted when not taken into account (Section 5.1).

(iii) **Which domain sources can be used to attract IPv6 scanners?** We conduct the most comprehensive study to date on the role of domain names for IPv6 scan attraction. We expose domain names through three different public domain sources (7 TLD Zone files, CT logs, as well as the Tranco top list) and compare exposure strategies by their effectiveness at attracting IPv6 scanners, all while mitigating the unwanted side effects of Hitlist inclusion (Section 5.3).

(iv) **What influence does the IPv6 Hitlist have on IPv6 scan traffic?** We analyze and quantify the extent to which IPv6 hitlists act as a mediating factor when studying how IPv6 scan traffic is attracted. We find that the IPv6 Hitlist [14, 55] has a substantial effect on IPv6 scan traffic received, which might have impacted previous work (Section 6).

Based on our findings, we compile a set of actionable recommendations for researchers trying to set up their own IPv6 telescopes (Section 7.1), and—to further allow others to build on our work—we release both the source code of our DNS-aware IPv6 telescope as well as the scan packets we captured as an artifact (Appendix B).

2 Background: Discovering Domains and IPv6 Addresses

A complete scan of the IPv6 address space is currently infeasible, given its immense size of 2^{128} ($\sim 3.4 \times 10^{38}$) unique addresses. As an optimization, IPv6 scanners first search for (likely) in-use

IPv6 addresses to select those addresses as scan targets. Scanners will combine different sources of IPv6 addresses in an attempt to achieve maximal coverage of the in-use IPv6 address space.

2.1 Domain Names

One way of collecting IPv6 addresses is through *domain names*. The intuition is that, if a DNS name has an `AAAA` record containing an IPv6 address, that address is more likely used by an actual live host. As a preliminary step, scanners therefore need to discover lists of domain names for which they then resolve the DNS records. In this subsection, we explain the three domain sources which we use to expose or telescope addresses.

Centralized Zone Data Service (CZDS). Zone files list all domains within a given DNS zone. Within the context of domain discovery, this typically concerns TLD zone files, which list all Second Level Domains (SLDs) under a specific TLD. With the delegation of new generic TLDs (gTLDs), the Internet Corporation for Assigned Names and Numbers (ICANN) introduced the CZDS to help interested parties get access to TLD zone files. The CZDS is commonly used by researchers or scanners to collect up-to-date lists of domains for further evaluations. This access is granted at a registry's discretion based on individual requests. Registries have to update the zone file of each gTLD daily. Next to new gTLDs, previously established gTLDs such as `.com` or `.net` are also published via this service. However, the CZDS does not contain any country code TLD (ccTLD). Certain ccTLD registries make their zone files available, either as open zone files that can be accessed publicly (e.g., `.ch`) or as closed zone files that a registry may selectively grant access to (e.g., `.nl`). However, most ccTLDs do not provide any external access to their zone files [44].

CT Logs. In the the Certificate Authority (CA) ecosystem, most CAs publicly log their issued certificates. This allows domain operators to monitor that CAs do not issue certificates for their domain unless authorized [24]. With CT becoming required by browsers, many CAs participate in CT logging, especially those in the web ecosystem. Since each logged certificate contains the domain(s) it is valid for and as logs can be monitored by anybody, CT logs became a source for domain names as a side effect. In comparison to CZDS, CT logs allow collecting subdomains besides the SLDs contained in TLD zone files, and CT logs cover large fractions of ccTLDs [44]. With domain operators often requesting certificates soon after creating a domain name [42], CT logs became a near-real-time source of new domain names. In addition to using full zone files, extracting domains from CT logs is also a common way to obtain domain names [13, 32].

Top Lists. Top lists enumerate domain names that are the most popular according to a chosen metric (e.g., website or DNS traffic volume) as identified from the list's data source (e.g., browser telemetry or a DNS resolver). Top lists provide a direct source of ranked, well-known domain names, and are used for various purposes such as marketing (indicating consumer attraction) or security (attaching reputability and trustworthiness to regularly visited domains). They are usually limited to one million domains and are updated once a day, though Google's Chrome User Experience Report (CrUX) list is an exception to both, listing tens of millions of domains and being updated once a month. Researchers and scanners commonly use top lists to get an up-to-date sample of well-known domains that they can analyze in their scans and experiments [33, 36, 40].

2.2 Hitlists

Besides domain names, other sources of IPv6 addresses exist, such as traceroutes, where IPv6 addresses are collected at each hop, or client addresses collected at Network Time Protocol (NTP) servers [37]. Given the variety of sources, centralized IPv6 address collection services, also known as hitlists, have emerged and have become a common source of IPv6 addresses. Hitlists group IPv6 addresses across sources and output them as a single list. While these services are useful as a one-stop shop for addresses, they may have a limited update frequency or may be subject to

Table 1. **Comparison of this study with related work that attracted scanners to IPv6 telescopes using domain names**—While existing studies relied on classic passive or reactive telescopes (the latter being telescopes that respond to incoming packets), our study introduces the concept of a *DNS-aware telescope* that allows for deeper insights into the conduct of domain-based IPv6 scanners. Note that Egloff et al. do not primarily study domain-based scanners—our comparison only pertains to one of multiple telescopes in their work. *Legend:* ✓ yes, ○ partially, × no.

Study		Domain exposure method: TLD	CT logs	Top list	rDNS	Considers Hitlist?	Telescope type
Tanveer et al.	[49]	`.com`, `.net`	×	×	×	×	passive telescope
Zhao et al.	[54]	×	×	×	✓	○*	passive and reactive telescopes
Egloff et al.	[11]	`.berlin`†	*unknown*	✓‡	*unknown*	○§	passive telescope
Tanveer et al.	[48]	`.com`, `.net`, `.org`	✓	×	×	✓	passive and reactive telescopes
This work		**`.com`, `.net`, `.org`, `.xyz`, `.ch`, `.nl`, `.de`**	✓	✓	×	✓	**passive *DNS-aware* telescope**

* Zhao et al. mention that some of the addresses from their experiments appeared on the public hitlist and that this could have affected their findings.
† We learned the TLD used by Egloff et al. in private correspondence with the authors as it is not specified in their paper.
‡ Egloff et al.'s domain is part of the Umbrella Top 1M top list.
§ Egloff et al. discuss the influence of the Hitlist regarding their experiments using BGP to attract scanners, but not regarding those using domain names.

access or usage restrictions. Scanners might therefore decide to continue collecting IPv6 addresses themselves, and we analyze this independent IPv6 address discovery in this paper.

A commonly used hitlist is the TUM IPv6 Hitlist service by Gasser et al. [14, 55]. It has been collecting addresses from different sources since 2018. These sources include: *(i)* addresses from DNS scans of domains retrieved from top lists, CZDS, static ccTLD lists and CT logs; *(ii)* addresses from different traceroute campaigns (e.g., RIPE Atlas [35], CAIDA Ark [3]); *(iii)* addresses from external data providers like Rapid7 and IPinfo.

Addresses collected from one of the data sources form a cumulative *input list* that is published after each run of the service. Besides collecting these raw addresses, the service further identifies and filters out specific addresses and then tests if the remaining addresses respond to probing packets. The Hitlist service sends a set of probes to check if the addresses respond to various protocols, i.e., ICMP, TCP port 80 and 443, and UDP port 53 and 443. Next to the input list, the service then publishes individual output files of addresses responsive to specific protocols as well as a combined list of all responsive addresses. The latest snapshot of the combined list of responsive addresses can be accessed freely, while historic data, input addresses, and the individual response files are accessible after an access request.

The IPv6 Hitlist operators allowed us to run a controlled experiment to evaluate the impact of the Hitlist on IPv6 telescopes (see Section 4.4).

3 Related Work

Network Telescopes are a common tool in Internet measurement research to evaluate traffic such as Internet-wide scans or backscatter [1, 5, 9, 12, 16, 18, 28, 29, 31, 43, 51]. Most studies to date focus on IPv4 because the small address space allows one to easily observe scan events at telescopes. In IPv6 on the other hand, the research method has shifted from merely observing telescope traffic to actively exposing telescope addresses and prefixes to IPv6 scanners, attracting scanners by giving the impression that the network contains actual live hosts worth targeting.

Tanveer et al. [49] were the first to construct an IPv6 telescope that attracts scanners by signaling "liveness." They tested various ways to expose IPv6 addresses from their telescope (e.g., sending web requests, registering a domain, participating in an NTP pool) and quantified their respective

effect on the scan traffic received. In contrast to our study, their work only focuses on scanners targeting the neighboring address space of the IPv6 addresses they exposed, owing to constraints in their experiment design.

Building on the work of Tanveer et al., further studies have experimented with different address exposure methods (Table 1). Zhao et al. [54] exposed IPv6 addresses via various reverse DNS (rDNS)-based techniques. Egloff et al. [11] exposed their telescope's address space to scanners using BGP announcements, while also collecting traffic in a partially active prefix with a domain name pointing to it. Klopsch et al. [22] studied scanners that collect IPv6 client addresses from NTP servers for scans, but found only two actors using this technique.

A recent follow-up study by Tanveer et al. [48] built a large network telescope that incorporates various exposure methods, including domain names exposed via zone files and TLS certificates. In addition to domain names, Tanveer et al. [48] also attracted scanners using BGP announcements and by placing domains on the IPv6 Hitlist. This study was the first to conclusively show that the DNS-based IPv6 scanners it attracted were different from those using the IPv6 Hitlist.

In our study, we present a new kind of telescope that allows us to analyze previously unexplored dimensions of domain-based IPv6 scanners. Through our *DNS-aware telescope*, we can uncover details to which previous work has remained oblivious: we can measure how quickly DNS-based IPv6 scanners scan the addresses they collect, as well as which scan sources share the same DNS-resolution infrastructure. This allows us to *(i)* show that IPv6 scanners with small vs. large scan delays are fundamentally different in their behavior and to *(ii)* identify distributed scanners that operate from multiple ASes. Our study also includes the largest comparison of different domain exposure methods to date, exposing domain names in seven different TLDs as well as CT logs and the Tranco top list.

As explained in Section 2, the TUM IPv6 Hitlist [14, 55] is a common source of addresses for IPv6 scans. Some of the previously discussed studies acknowledge that their exposed addresses end up on the IPv6 Hitlist [11, 54]. It is, however, often unclear to what extent scan traffic attracted to telescopes can be attributed to the actual exposure method as opposed to the IPv6 Hitlist—especially in cases where the attraction method involves DNS records. In our study, we therefore collaborate with the operators of the Hitlist to *(i)* exclude our telescope IP ranges from the IPv6 Hitlist, and *(ii)* we perform additional experiments to quantify the effect of Hitlists on IPv6 scan traffic. This is thus the first study that investigates domain-based IPv6 scanners in complete isolation, without any mediating effects through the IPv6 Hitlist.

Previous works have studied scanners that rely on publicly available domain names, without explicitly focusing on IPv6 [23, 26, 32, 39]. For example, Pletinckx et al. [32] evaluate whether scanners use domains exposed via certificates in CT logs and how fast they respond to newly published certificates. Unfortunately, these studies are either IPv4-only, or they do not take into account that some of the IPv6 scanners observed were possibly attracted through the IPv6 Hitlist.

Finally, Richter et al. [34] studied IPv6 scan traffic not through a traditional network telescope, but through firewall logs at a large CDN. As the firewall only logs scans to ports other than port 443 and 80, their research was limited to scanners that do not target web services. In our study, we show that a considerable amount of IPv6 scanners do in fact target ports 80, 443, which Richter et al. were unable to observe.

4 Method

An overview of our method can be seen in Figure 1. It can be divided into four parts:

(i) we set up a /36 network telescope capturing all incoming IPv6 traffic and an IPv4 web server (Section 4.1);

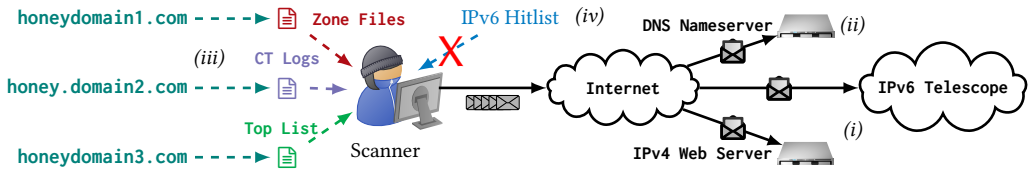


Fig. 1. **Overview of our method**—*(i)* We set up an IPv6 telescope and an IPv4 web server (see Section 5.5). *(ii)* We register different domains and set up an authoritative nameserver. *(iii)* We expose the registered domains in different sources. *(iv)* We exclude our telescope's address space from the IPv6 Hitlist.

(ii) we register 62 *honey domain names* and configure a nameserver to respond with unique AAAA records to addresses in the telescope's address space for each request (Section 4.2);
(iii) we expose the honey domain names in three publicly accessible domain sources that we suspect scanners might be using (Section 4.3); and
(iv) in collaboration with the IPv6 Hitlist operators, we exclude our telescope from the IPv6 Hitlist and publish a set of control addresses from a dedicated /56 telescope in different parts of the Hitlist (Section 4.4).

4.1 Telescope Setup

Our telescope consists of a previously unused /36 IPv6 prefix, which is announced in BGP as part of a larger /32 covering prefix. Although the specific prefix of the telescope is not announced individually, the telescope might still receive probing packets from unrelated IPv6 scanners that, for example, discovered active hosts in the neighboring prefixes or observed the BGP announcement for the covering prefix. To account for such background traffic unrelated to our experiments, we split the /36 prefix of our telescope in two: *(i)* we use the addresses in the upper /37 as honey addresses, which we expose through domain names; and *(ii)* we leave the lower /37 untouched to assess how much baseline traffic the address space receives over time.

We choose to operate our prefix as a passive telescope, i.e., we do not respond to any incoming packets. This way, we prevent the addresses in our telescope from being indexed by Internet intelligence platforms like Censys or Shodan, which would attract additional scanning traffic unrelated to our exposure methods.

In addition to our IPv6 telescope, we set up an IPv4 web server. We use this web server in Section 5.5, to evaluate whether scanners target IPv6 specifically or focus on the domain independent of the IP version, e.g., by preferring IPv6 due to Happy Eyeballs [41], but falling back to IPv4 if necessary. Given the scarcity of IPv4 addresses, the web server is configured with a single address.

4.2 DNS Setup for Honey Domains

Scanners that want to use domains to scan the IPv6 address space need to perform two steps in their target reconnaissance process: First, they need to resolve DNS AAAA records containing one or more IPv6 addresses and, subsequently, scan the IPv6 address(es). This can result in two different types of scanning behavior: on one hand, a scanner can resolve the domain and perform the following IPv6 scan immediately afterwards, resulting in small delays between the DNS request and the IPv6 scan probes. On the other hand, large-scale IPv6 scanners might separate the preliminary address collection phase (i.e., querying AAAA records for many domains in bulk) from the subsequent scan phase (i.e., sending IPv6 probe packets to all addresses collected). In this case, there will be a greater time delay between its DNS queries and when it starts sending actual IPv6 scan packets.

To differentiate these two groups, we need to map DNS requests to scans. We therefore designed and developed a custom DNS server setup that allows us to distinguish between the two scanner

types. We configure the authoritative DNS server of all honey domains to hand out two *unique* IPv6 addresses from the /37 telescope for each AAAA query it receives. Additionally, we set the time-to-live (TTL) value in all DNS responses to 1 to prevent resolvers from caching our domains' AAAA records.[1] As the DNS server hands out different addresses for each query, we can trace back packets scanning exposed addresses in the telescope to the DNS answer which resolved the domain. We can hence measure the elapsed time Δt between the point at which the DNS server first handed out a given IPv6 address and the point at which a scanner scanned that address in the telescope.

4.3 Exposure of Honey Domain Names

To cover a variety of sources that scanners might be using to discover domains (Section 2.1), we expose our 62 honey domain names in three types of sources:

TLD Zone Files: We register 28 domains under seven different TLDs to attract scanners through the respective TLD zone file. Four of the TLDs, .com, .net, .org, and .xyz, have zone files that are available via CZDS. In addition to these, we use one ccTLD with an open zone file (.ch), and two ccTLD with closed zone files (.nl and .de). For every TLD, we register four domains each.

CT Log: We obtain Transport Layer Security (TLS) certificates for 8 subdomains under 4 dedicated SLDs ($2 \times$ xyz, $2 \times$ org), thereby exposing them to scanners through CT Logs. By using subdomains, we ensure that the domains we expose are not part of any published TLD zone file but only visible in CT Logs. Additionally, we also register wildcard certificates for 8 subdomains and configure our DNS server to issue responses for every possible domain that matches the wildcard. This allows us to study if scanners try to guess subdomains for wildcard certificates in CT logs. We renew our Certificates every week such that our domains keep appearing in CT logs over time.

Tranco Top List: We collaborate with the maintainers of the Tranco top list to temporarily place 18 honey domains ($9 \times$ xyz, $9 \times$ org) on the list. As the main Tranco top list only contains second-level domains, these domains will also attract scanners through their TLD Zone files. Since we do not want to misidentify these as scans attracted through Tranco, we rigorously filter these scans based on our other TLD zone file experiments: we only attribute a scan to the Tranco top list if it occurred *after* we included our domains in Tranco, and if we did not receive *any* scan packets from the same source AS to any of our other domain names exposed via the respective TLD. As the Tranco top list is also used by researchers, we take precautions to prevent our domain names from having a statistically significant effect on studies using Tranco, which we discuss in Appendix A. Two domains are within the lower-ranked half of the Top 10k, 12 domains within the lower half of the top 100k, and 12 domains within the lower half of the Top 1M.

4.4 IPv6 Hitlist Experiments

As explained in Section 2, the IPv6 Hitlist service [14, 55] is a commonly used tool by scanners and publishes different lists of addresses. Therefore, we want to address two important questions in this regard. *(i)* How substantial is the effect of the IPv6 Hitlist service on IPv6 scan traffic? *(ii)* How prevalent is domain-based IPv6 scanning if one excludes scanners attracted through the Hitlist?

To prevent scanners from finding our telescope through the IPv6 Hitlist, we collaborate with the Hitlist operators and explicitly exclude all addresses from our /36 telescope from the Hitlist. Thus all scanners have to resolve our domains themselves or use other sources to discover our telescope.

Furthermore, we perform additional experiments to quantify the effect of the Hitlist on IPv6 scan traffic. We set up a dedicated /56 telescope and place specific addresses from the telescope on each of the published address lists, i.e., the combined list of responsive addresses (openly accessible, labelled *public*), and the cumulative input (labelled *input*), as well as lists specific for the five probed protocols (*icmp, tcp80, tcp443, udp53, udp443*, accessible after registration). To each list we add two

[1]We choose a TTL of 1 over 0 as the latter has been known to cause issues with some DNS resolvers [27].

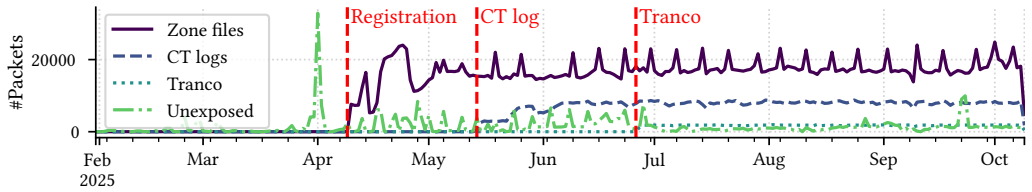


Fig. 2. **Daily received IPv6 packets at the telescope over time across different domain exposure methods**—To improve readability, the figure excludes data from AS139831 because it sent 643.8 k probes on a single day. The "unexposed" category shows traffic to addresses that were not exposed via DNS. The red lines denote the dates at which we first exposed domains in Zone files, CT logs, and Tranco, respectively.

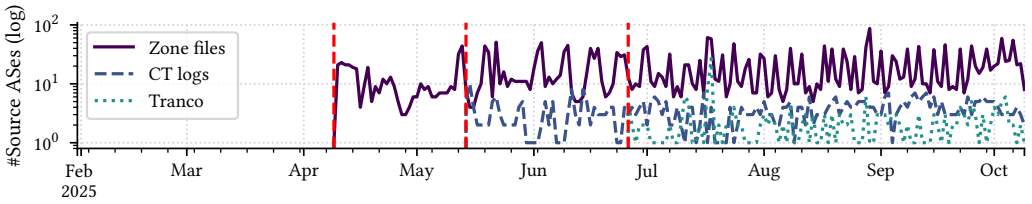


Fig. 3. **Daily number of source ASes at the telescope over time across different domain exposure methods**—Note the logarithmic y-axis. To improve readability, we do not show scans to unexposed addresses.

random addresses, two EUI-64 addresses, two low-byte addresses and two addresses encoding IPv4 addresses in the last 4 bytes respectively, resulting in eight addresses published in each of the seven lists, totalling 56 published addresses. These addresses are not published via any other channel, i.e., they can only be found via the Hitlist. This allows us to evaluate traffic attracted by the IPv6 Hitlist compared to the domains we expose. Adding different addresses to different sources allows us to evaluate which lists are commonly used (e.g., do scanners focus on only responsive addresses or all known addresses from the input?).

5 Analysis of Domain-Based IPv6 Scanners

Using our method for attracting IPv6 scanners and tracing their traffic back to our exposed honey domain names, we now analyze this traffic to identify how domain-based IPv6 scanners behave.

We first registered all our domains on April 14, 2025 and collected traffic until October 9, 2025. Starting on May 14, we requested TLS certificates for a designated set of (sub)domains every week. We included our domains in the Tranco Top 1M for four days starting on June 27.

Over our total observation period, our telescope received 9.4 M packets from 25.5 k addresses originating from 613 ASes. Out of these, 8.4 M packets (89%) from 25 k addresses (98%) originating from 592 ASes (97%) targeted the addresses that were explicitly handed out by our DNS server. Figure 2 shows the number of scan packets received at our telescope over time. We started captures at the telescope before exposing any domain. As seen in the figure, traffic is mostly zero during this period, but some probes found their way to the telescope, most notably on April 1, where a well-known research group [19] sent 32.8k packets from AS1764/NextLayer. After we expose our domains, the telescope collects thousands of probe packets per day. A major traffic spike occurred on October 6, where a single AS—AS139831/DITO—sent 643.8 k probes (we exclude this AS from Figure 2 to avoid stretching the proportions).

For scanners scanning zone files, we see a clear weekly pattern with an increased number of incoming probes on Mondays. This is attributable to a specific set of source addresses in AS14061, which we cover in detail in Section 5.1.

Table 2. **Top 10 ASes based on the number of packets sent towards our telescope**—The "unexposed" column indicates if the AS also scanned addresses that were not handed out by the DNS server.

Rank	ASN	AS org.	#Packets↓	#Src. addrs.	#Dst. addrs.	Unexposed?	Domain sources used	Protocols	#Ports	#Days active
#1	AS211298	Driftnet	7 120 907	258	1570	✗	`.com`, `.net`, `.org`, `.xyz`, CT log	TCP, UDP	43 918	182
#2	AS139831	DITO	643 820	2	643 806	✓	`.org`, `.xyz`	TCP, ICMPv6	1	2
#3	AS13335	Cloudflare	445 645	14 289	52 554	✗	`.com`, `.net`, `.org`, `.xyz`, `.ch`, `.nl`	TCP	2	183
#4	AS6510	BYU	371 801	2	3676	✗	Tranco, `.xyz`	TCP, ICMPv6	2	105
#5	AS14061	DigitalOcean	286 884	4136	6283	✓	`.com`, `.net`, `.org`, `.xyz`, `.ch`, `.nl`, CT log, CT wildcard	TCP, UDP, ICMPv6	317	162
#6	AS23910	CERNET2	97 548	13	86 071	✓	–[1]	TCP, ICMPv6	17	94
#7	AS38272	CERNET	85 725	3	85 725	✓	–	ICMPv6	0	101
#8	AS132203	Tencent	83 130	6	42 120	✓	–	ICMPv6	0	43
#9	AS1764	Next Layer	32 771	1	4096	✓	–	ICMPv6	0	1
#10	AS51224	link-lab	28 145	1	11 726	✓	–	ICMPv6	0	10

[1] One address in AS23910 sent four TCP packets to an address that was resolved from one of our Tranco domains, albeit before the domain was exposed on the top list. We presume that the scanner found the domain in the `.xyz` zone file.

Next, we turn our attention to the source ASes of the scans.[2] Figure 3 shows the daily number of ASes scanning the telescope. While the number of ASes fluctuates, there are events that stand out: On July 18, the number of ASes targeting the Tranco domains briefly increases to 27, whereas this normally is below 10. On August 29, an even larger spike occurred, as the telescope saw a sharp uptick in scanners targeting `.xyz` domains (85 ASes on a single day). We take advantage of our one-address-per-query method to investigate if these scanners might be related. While the Tranco scanners of July 18 all target different addresses, we find that 80 ASes of the `.xyz` scanners of August 29 targeted a *single address*, even though every DNS query for our honey domain names returns a unique address. The address in question was handed out by our DNS server 12 days earlier in response to a query from an academic network. According to rDNS records, several of the source addresses participating in this event belong to CAIDA Ark [3], a distributed measurement platform. Our novel method therefore allows us to determine that there was a delay between obtaining the IPv6 address from the DNS and scanning the address, and, additionally, that the address was shared between a large number of sources. We conclude that these `.xyz` scanners are therefore most likely part of a distributed academic scan.

5.1 Top Scan Sources

In this analysis we investigate the top scanning ASes. Table 2 shows the top ten source ASes in our telescope by the number of packets sent. The highest-ranked AS is an acknowledged scanner[3] that belongs to *Driftnet*, a British cybersecurity company. They target domains from zone files and CT logs, and run broad port scans to more than 43.9k ports across our entire measurement period, resulting in over 7 million packets observed. In contrast, over 99% of traffic from the second-highest AS (AS139831–DITO) consists of ICMP probes originating from a single address, which probes several unexposed addresses throughout the telescope. This AS also contained a second address that sent 22 TCP packets to our `.org` and `.xyz` domains.

The two following ASes—Cloudflare and Brigham Young University (BYU)—almost exclusively send HTTP(S) probes (ports 80 and 443) to the domain addresses. In the case of BYU, this traffic came from a single address targeting Tranco domains.[4] The BYU AS also contains a less prolific

[2] As IPv6 address space is cheap, it is common for IPv6 scanners to distribute large-scale IPv6 scans across multiple source addresses in the same network [11, 34, 48]. To not overestimate the number of distinct entities responsible for scans, we quantify scanners by the number of unique source ASes whenever possible. This gives us a conservative estimate of the number of scanners that does not overcount single scanners operating from large prefixes.

[3] Acknowledged scanners are benevolent scanners that declare their behavior and purpose, e.g. on a public website [5].

[4] We reached out to the research group at BYU operating these scans and learned that this is an experiment on SYN Cookie adoption: each day, the researchers collect IP addresses from the Tranco Top 1M domains and add them to an aggregate list

address that participated in the on-off CAIDA Ark scan mentioned at the beginning of this section. Similarly, we can identify two sets of IP addresses in the Cloudflare AS. The majority of traffic (>99%) comes from over 13 000 addresses across the `2400:cb00/32` prefix. The rest originates from 868 addresses, which—according to the IPinfo privacy classification dataset [21]—are relay exit nodes from Cloudflare WARP [4].

DigitalOcean (#5) was another prominent source of scan traffic. Given that this is a hosting provider, it is unclear how many distinct entities are responsible for the traffic from this AS. Using rDNS records, we were able to attribute 378 addresses to BinaryEdge, and 77 to LeakIX—both cybersecurity companies. For the remaining addresses, we again leverage our one-address-per-query technique to estimate which source addresses within the telescope may be related. By grouping together source addresses that target the same destinations, we find 196 addresses that scan the same telescope addresses as BinaryEdge; these are likely addresses that belong to the same organization, but were missing the identifying rDNS records. We also find five groups of 574 to 851 addresses that share at least one destination address. Interestingly, while none of the addresses within these groups target more than 35 unique ports individually, the combined addresses in each group target the exact same 165 ports. When evaluating the traffic of the five groups longitudinally, we find that scanning patterns of the individual groups align: each group of IP addresses performs scans every Monday for three to five weeks before it stops scanning, and another group takes over. We conclude that these addresses are operated by the same entity, which distributes scans across multiple source addresses and swaps out its source address pool after a few weeks (see Figure 9 in Appendix C for a visualization). This shows that, using our method, we can also identify scanning campaigns across multiple addresses *within* an AS.

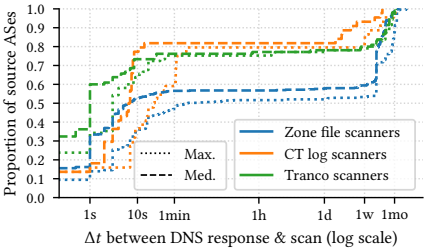


Fig. 4. **ECDF plot of the proportion of IPv6 source ASes vis-à-vis their maximum and median Δt (rounded to seconds)**—Across all exposure methods, the majority of scanners predominantly perform *early* scans, where the scanner scans an address immediately after receiving a response from the DNS server ($\Delta t < 1h$).

The remaining five ASes solely focus on addresses which our DNS server did not expose. These ASes predominantly send ICMPv6 traffic and (except #8) are either research ASes (#6, #7) or can be attributed to well-known research groups in the measurement community (#9 [19], #10 [11]).

Key Takeaways: (i) *The heavy hitters in our telescope comprise both unknown actors as well as known cybersecurity or research organizations.* (ii) *Using our method, we can check if addresses in cloud ASes are part of a common scanning campaign, as shown by the example of DigitalOcean.*

5.2 Early vs. Late Scanners

As explained in Section 4.2, we want to explore if domain-based IPv6 scanners resolve a domain and scan the corresponding address in one go, or if they have a dedicated address collection phase (i.e., they resolve the domain, collect the IPv6 addresses, and perform the scans at a later time). To investigate this, we focus on the time difference between DNS queries and scans. For every source AS scanning an exposed address in the telescope, Figure 4 shows the maximum and median time Δt between when the DNS server first handed out an address and when the AS scans the address in the telescope (split up by the exposure method of the associated domain name). For each domain exposure method, over fifty percent of the ASes always scan the telescope in less than a minute

of addresses. They then routinely perform a measurement to the addresses on this lists that sends multiple TCP packets to ports 80 and 443, explaining the high amount of traffic we observe.

after receiving an answer from the DNS server. On the other end, we see a long tail of ASes that send scan packets to the address weeks or even months after it was first handed out. The AS with the largest Δt of 53 days was AS1149. According to WHOIS data, this AS was used for research scans by the University of Twente at the time. In the ASes scanning CT log and Zone File domains, we see two spikes after 7 days and 12 days, respectively. In case of the latter, this can be attributed to the 80 ASes in the distributed scanning event mentioned in Section 5. The former are three ASes from the Censys cybersecurity company (AS398324, AS398705, AS398722), which scan the CT log domains almost exactly one week after receiving the IP address from the DNS server.

Our findings show that domain-based IPv6 scanners fall into two categories: *early scanners*, which scan the address shortly after resolving the domain name, and *late scanners*, which resolve the domain name but then take time before sending any IPv6 scan packets. In the remaining sections, we will refer to scans with $\Delta t < 1h$ as *early* scans and to those with $\Delta t > 1h$ as *late* scans.

Key Takeaways: (i) *We find that domain-based IPv6 scanners can be classified as early or late depending on their address collection and scanning procedure.* (ii) *Across all attraction methods, over 50% of the scanners perform only early scans.*

5.3 Which Domain Sources Do Scanners Use?

We now turn to the question of which of our domain exposure methods were most effective at attracting IPv6 scanners. Table 3 shows both the amount of scan traffic as well as the number of sources we attracted using different domain exposure methods. The zone-file-based exposure method attracted by far the most scanners (16 777 unique source addresses in 488 ASes). gTLDs were most effective (13 982 unique source addresses in 462 ASes total), followed by the `.ch` ccTLD (7003 source addresses in 171 ASes). The `.nl` ccTLD attracted a moderate number of scanners (2046 addresses in 31 ASes), whereas the `.de` TLD only received traffic from three source addresses. This distribution reflects the open (gTLDs, `.ch`) versus closed (`.nl`, `.de`) nature of these zone files, directly correlating with how easy it is for organizations to discover these domain names.

Table 3. **Sources and traffic attracted to the IPv6 Telescope by various domain exposure methods**—gTLDs were the most effective attractor. CT logs, Tranco, and the `.ch` TLD also drew in many scanners.

	#Scans		#Sources		
Source	Early	Late	ASes	Addrs.	Dst. addrs
com	49 984	690 492	232	8186	8311
net	64 077	680 609	242	7899	12 947
org	46 747	721 589	203	7499	6921
xyz	55 849	715 868	297	7738	11 157
ch	45 690	31 306	171	7003	6738
nl	464	8500	31	2046	160
de	124	0	2	3	8
CT log	6805	1 095 484	44	3182	1380
CT wildcard	16	12 775	6	2858	90
Tranco	1609	185 851	105	119	2383

The subdomains for which we registered certificates received more scan packets than any other individual exposure method (1 095 484 packets). However, 99% of these originate from a single AS (AS211298/Driftnet; cf. Table 2). We also see six ASes trying to guess subdomains under the domains for which we registered wildcard certificates. Finally, the Tranco top list attracts several scanners as well (119 source addresses from 105 ASes).

In light of these findings, we ask ourselves: **are scanners we attract simply using multiple domain sources, or do different exposure methods actually attract different scanners?** To answer this question, we look at the intersections of source ASes across exposure methods.

Figure 5 shows an UpSet plot [25] of the largest subsets of source ASes attracted through different exposure methods. We see that there are substantial overlaps in the source ASes we attract through zone files, as seen in the rightmost portion of Figure 5. For example, 41 ASes send traffic to all of the five most-used TLDs in our experiments (`.xyz`, `.net`, `.com`, `.org`, and `.ch`), whose zone files are easily accessible. In addition to smaller intersections across different TLD zone files, we also see overlaps between scanners using zone files and those using Tranco or CT logs.

Apart from such multi-source scanners, we also observe several scanners that only target domains from a single source, as seen in the left half of Figure 5. Among these, most are targeting the .xyz (95 ASes) and Tranco (74 ASes) domains. In case of the former, this is mostly because of the 80 ASes from the singular scanning event targeting .xyz domains. If we remove scanners that only target the shared address during the distributed CAIDA Ark event from the .xyz scanner set, only 31 ASes remain. This is, however, not the case for the Tranco scan event mentioned earlier: if we only include scanners that sent packets on days other than one where we saw the Tranco scan event, the number of ASes in the Tranco-only set merely drops to 45.

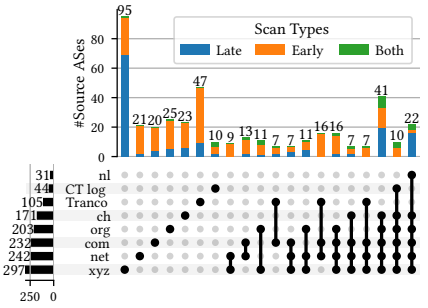


Fig. 5. **UpSet plot showing the intersections of source ASes across different exposure methods**—Hues indicate how many ASes in a set are performing only early scans (orange), only late scans (blue), or both early and late scans (green). Owing to space constraints, we only show subsets containing more than seven ASes, and we omit scan traffic targeting .de or wildcard certificate domains.

Among the .net, .com, .org, and .ch TLDs, we also see more than 19 ASes that each are only attracted using the respective zone file. This shows that, even though there are overlaps between scan sources, it is worthwhile for IPv6 telescope operators to expose domains in multiple sources as there are several scanners that only target domains in specific sources.

Key Takeaways: (i) *gTLDs were the most effective attractor. CT logs, top lists, and the* .ch *TLD also drew in a considerable number of scanners.* (ii) *While there are substantial overlaps in the source ASes attracted using different domain sources—especially across different zone files—several scanners are only attracted using a single domain source, which makes it worthwhile for telescope operators to expose domains in multiple places.*

5.4 Differences in Scanner Behavior

The two categories of early and late scanners not only reflect different strategies for address collection: they also correlate with different *scanner behavior*. In Figure 6, we plot the protocols scanned by ASes in early versus late scans (recall that "late" scans are scans where the packet arrives at an address >1h after the address was handed out by the DNS server, cf. Section 5.2). When performing late scans, most (i.e., 232 out of 257) ASes exclusively send ICMP packets. However, among the early scans there is no AS that *only* scans ICMP—here, ASes either scan TCP/UDP (371 ASes) or both ICMP and TCP/UDP (19 ASes).

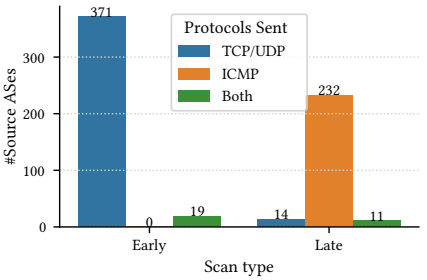


Fig. 6. **Protocols sent by source ASes in early vs. late scans**—Early and late scans differ greatly: early scanners almost exclusively send TCP/UDP traffic whereas late scanners focus on ICMP.

Targeted Ports and Services. Among those ASes also scanning TCP/UDP (390 for early scans and 25 for late scans), only 72 (20%) of the scanners that did early scans scanned ports *other* than 80 and 443. Apart from 80 and 443, the most common ports scanned in early scans are SSH (port 22; 58 ASes) and SMTP (ports 587 and 25; 15 and 6 ASes, respectively). The ASes doing late TCP/UDP scans also focus largely on web ports (ports 443 and 80; 18 and 14 ASes, respectively), closely followed by SSH (port 22; 10 ASes), and DNS (port 53; 9 ASes). Overall, we observe that late TCP/UDP scanners target substantially more ports than early ones: despite being fewer in number, the 25 late TCP/UDP source ASes target 44K ports in total, while the scans from early ASes cover only 18. (For a more in-depth analysis of this, see Appendix F.)

Differences in ICMP Scans. When it comes to ICMP scans, we also see substantial differences between early and late scanners. Figure 7 shows a CDF of the median hop limit of the ICMP packets we receive per source AS. For 58% of all the ASes doing late ICMP scans, the median hop limit in the ICMP packets is smaller than 6. This indicates that these scanners are most likely using the IP addresses as targets in *traceroute measurements*. When performing traceroutes, the scanners will start sending packets with low but incrementing hop limit values to trigger ICMP error messages from the routers along the path—as the destination is our telescope, we receive these probing packets with small hop limit values. For the majority of the remaining late ASes, the ICMP packets arrive with a median hop limit slightly below 255. These ASes also comprise the 80 ASes from the distributed CAIDA Ark scan. In the early scans, however, the ICMP packets of almost all ASes have a median hop limit slightly below 64. This shows that IPv6 scanners are using different ICMP scanning tools, which correlate with their initial address collection strategy.

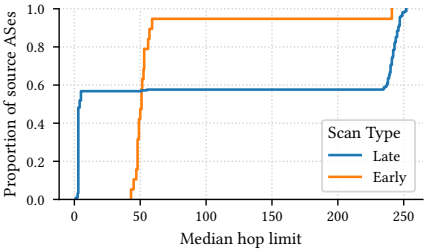


Fig. 7. ECDF of the median hop limit per source AS for ICMP packets in early vs. late scans.

Differences in AS Types. We also investigate the AS type of the source ASes observed in our telescope (Figure 8) using IPinfo's IP-to-ASN API [20]. While both early and late scans predominantly originate from hosting ASes, ISPs are a far more common source in early scans, and education ASes are comparatively more prevalent among late scans. The latter indicates that a substantial portion of IPv6 scanners with a dedicated scanning phase are likely research-related.

Key Takeaways: (i) *We find stark behavioral differences between early and late scanners, with early scanners scanning mostly TCP/UDP, while late scanners predominantly scan ICMP.* (ii) *Among the early scans, the overwhelming majority of source ASes* only *scan web-related ports (80 and 443).* (iii) *Based on the hop limit values, we observe that more than half of the source ASes in late ICMP scans are likely performing traceroutes.* (iv) *The source AS types also differ between late and early scanners.*

5.5 Are Early Scanners Dual-Stack?

In the previous subsection, we showed that early IPv6 scanners predominantly send TCP/UDP traffic to ports 80 and 443. Since these scanners are focusing on web protocols, the question arises: **are these scans specific to IPv6, or are they generic web crawls that could also be observed in IPv4?** It is possible that some of the scanners we observe are merely trying to crawl domain names without any regard for which underlying layer-3 protocol they are using. In theory, these scanners could be running a dual-stack setup where they merely connect via IPv6 on a best-effort basis and fall back to IPv4 if they are not able to establish a connection.

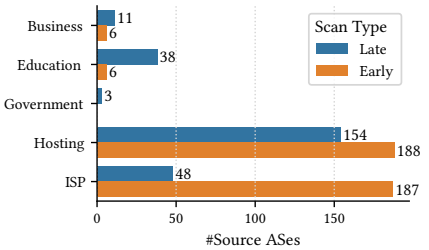


Fig. 8. AS categories of source ASes in early vs. late scans.

As introduced in Section 4, we set up an IPv4 web server under our control to evaluate this behavior. For every domain we use to attract IPv6 scanners, we configure an extra A record that resolves to the IPv4 address of the web server. As our telescope does not respond to probing packets, a dual-stack domain scanner might try to connect to our IPv4 web server once it receives no response via IPv6. By logging the incoming probes at the web server including their source addresses and arrival times, we can look for instances where a scanner tried to connect to both our IPv6 telescope and our IPv4 server.

Table 4. Top 6 HTTP requests of dual-stack crawlers among the early scanners by number of source ASes.

Rank	#ASes	Request
#1	80	`GET / HTTP/1.1`
#2	19	`POST /xmlrpc.php HTTP/1.1`
#3	13	`GET /.env HTTP/1.1`
#4	11	`HEAD /wordpress.zip HTTP/1.1`
#5	11	`GET /favicon.ico HTTP/1.1`
#6	10	`HEAD /public_html.zip HTTP/1.1`

Table 5. Scanning traffic received for addresses from each specific dataset of the IPv6 Hitlist.

	#Probes					#Sources	
Dataset	Total	TCP	UDP	ICMP	#Dst. Ports	ASes	Addrs.
`public`	384 807	364 449	16 421	1560	8566	37	5585
`input`	174 660	162 558	6783	2723	209	113	8386
`icmp`	12 841	9336	48	3457	759	84	148
`tcp443`	3250	3236	0	14	392	3	3
`tcp80`	3359	3237	0	122	392	4	12
`udp443`	101	87	0	14	8	2	2
`udp53`	889	299	576	14	9	5	5

As we do not know which IPv4–IPv6 address pairs belong to the same scanners, we match the scans at the IPv6 telescope to those at the IPv4 web server based on the source AS. If our web server observes a connection *(i)* from the same AS *(ii)* to the same domain name *(iii)* on the same destination port *(iv)* within 1 second of a corresponding connection attempt in the IPv6 telescope, we consider it a scan from a dual-stack scanner. This gives us a conservative estimate of how many scanners may be employing such a dual-stack scanning technique.

In the telescope, we observe 12.6 k addresses in 360 ASes that perform early scans of ports 443 and 80. The number of source addresses is deceptively large—10 k of these addresses are located in AS13335 (cf. Table 2). Across the 360 ASes, the median number of source addresses performing early web scans is 1 (only 157 ASes contain more than one early web scan address). If we match the scans of these addresses with the connections we see at the web server, we can identify a matching IPv4 connection attempt for 590 addresses in 123 ASes. For 81 of these ASes, we could find a matching request for every IP address that performed early web scans. This shows that a large portion of early domain-based IPv6 scanners are not targeting IPv6 per se; instead, these scanners are most likely crawlers trying to access the web services hosted under the domain names. As these scanners are running on a dual-stack setup, they do not have any special regard for the underlying layer-3 protocol and will fall back to IPv4 if IPv6 is not available.

Since our web server is able to capture the HTTP request of scanners, we can use this information to gain insights into which web services IPv6 scanners/crawlers are targeting. Table 4 shows the top 6 HTTP requests at the web server that we were able to match to scans in the telescope. While the majority of scanners are merely sending a simple `GET` request for the web root, we also observe several ASes targeting resources that may be related to vulnerabilities such as WordPress' XML-RPC endpoint or `.env`. This shows that some web-protocol IPv6 scanners may actually harbor malicious intentions and thus pose genuine security threats to operators.

Key Takeaways: (i) *We show that a substantial portion of early IPv6 scanners are* layer-3 agnostic*: these scanners are performing web crawls on a dual-stack setup and will fall back to IPv4 if necessary.* (ii) *By cross-referencing HTTP requests at our IPv4 server with IPv6 scans in our telescope, we find evidence of scanners searching for vulnerabilities in web services.*

6 Influence of the TUM IPv6 Hitlist on IPv6 Scan Traffic

Having studied the impact of domain names on IPv6 scan traffic, we turn to the question: **would we have seen additional scan traffic had we not excluded our addresses from the IPv6 Hitlist (Section 4.4)?** This topic is especially important for future studies seeking to study IPv6 scanners, as researchers might risk conflating scanners attracted by their address exposure methods with those attracted by the Hitlist.

We set out to answer this question by exposing a total of 56 addresses in the cumulative input list, the joint list of responsive addresses, and the protocol-specific lists of responsive addresses of the IPv6 Hitlist Service (cf. Section 4.4). We collect the resulting traffic at an independent /56 network telescope.

Indeed, after collecting traffic for 200 days, our network telescope received a total of 587 977 probes sent to the 56 addresses published on the IPv6 Hitlist, originating from 11 k addresses in 213 ASes. 96.67 % of all probes use TCP (92.61 %) or UDP (4.05 %), scanning a total of 8864 different ports (8257 TCP, 741 UDP), whereas only 2.48 % are ICMPv6 echo requests; the remaining 0.85 % are Generic Routing Encapsulation (GRE) packets or ICMPv6 echo replies. Of the source addresses we observe targeting addresses exposed on the Hitlist, 7244 (66%) in 183 ASes (86%) also scan one of the exposed addresses in our domain telescope.

We also investigate which of the datasets published by the IPv6 Hitlist attract the most traffic. As seen in Table 5, the public list of all responsive addresses (labeled `public`) attracts the highest number of scanning probes, which mainly consists of TCP-based traffic targeting a wide range of ports from 37 ASes. While the `input` dataset also attracts many TCP probes, they originate from a wider range of ASes and target fewer ports. Note that the Hitlist input contains all addresses collected by the Hitlist, regardless of whether they are responsive. This shows that merely having an address collected by the Hitlist will increase the number of observed IPv6 scan sources considerably, even if this address is located in an unresponsive IPv6 telescope.

The protocol-specific lists of responsive addresses attract less traffic: the `icmp` dataset attracts the most ICMP echo requests as well as a large number of probes targeting various TCP ports, coming from a total of 84 source ASes. The lists of addresses responsive on HTTP port numbers (`tcp80`, `tcp443`, `udp443`) see mostly TCP probes targeting non-standard HTTP ports (8899, 8090) but also ports for other protocols (22, 554). Scanners using the `udp53` dataset send 89 % of their probes towards port 53, using 73 % UDP and 27 % TCP. In Appendix E we provide additional analyses on the scanning traffic we observed at our Hitlist addresses.

Key Takeaways: (i) *Having an address in the IPv6 Hitlist attracts a substantial amount of scan traffic from multiple ASes. Future studies should therefore carefully consider if this "Hitlist effect" could impact their findings.* (ii) *Even non-responsive addresses can attract scanners through the Hitlist if they appear in the cumulative input.*

7 Discussion

7.1 Five Lessons for Building Domain-Based IPv6 Telescopes

In this section, we revisit our key findings and discuss what researchers operating IPv6 telescopes can learn from them to study IPv6 scanners more effectively.

(i) **Beware of the Hitlist Effect.** A central question throughout the existing literature on building IPv6 telescopes has been "how can we expose IPv6 addresses to attract IPv6 scanners?" Researchers commonly try to answer this by first exposing addresses and prefixes from their telescope in various places, checking if there are any incoming IPv6 scan packets, and then assessing if the exposure method is an effective way to attract scan traffic. As we have shown, there is a common fallacy here: if the address in question ends up on the IPv6 Hitlist, it will attract large amounts of IPv6 scan traffic either way. To study how scanners find target addresses, one should account for the IPv6 Hitlist, either by (1) asking hitlist operators to exclude the telescope's address range or by (2) having an adequate control to measure the base rate of scan traffic attracted through the Hitlist.

(ii) **Check for Distributed Scanners.** IPv6 address space is cheap. It is therefore easy for scanners to distribute their scans across different source addresses, either to send packets faster by using multiple scanning instances or to evade rate limiting or blocking at the target. Consequently, it is hard to tell if different scan sources—within the same or different ASes—belong to the same entity. Our DNS-aware IPv6 telescope can uncover such related scanners by mapping scan traffic from different sources back to a shared initial DNS query. Researchers interested in identifying distributed scanning efforts themselves can build on our work as we publish our source code and instructions on how to replicate our custom DNS-setup as an artifact (cf. Appendix B).

(iii) **What Kind of Scanners Am I Studying?—Is My Traffic IPv6-specific?** We have shown that there are two fundamentally different types of domain-based scanners in IPv6. On the one hand, there are the *late* scanners: sophisticated dual-phase scanners that first collect IPv6 addresses and then later conduct large-scale scans, often from different source addresses and sending mostly ICMP probes (cf. Section 5.2). On the other hand, there are the *early* scanners that scan IPv6, but do not have a preliminary address collection phase. As we have shown, early scanners are often not confined to IPv6 and also fall back to IPv4 if they cannot establish an IPv6 connection. To put things another way: we discovered that early scanners are largely *service oriented* as they mostly send TCP/UDP probes to the ports of common layer 7 protocols, whereas late scanners are more topology or *reachability oriented* since they predominantly send ICMP packets—either as a responsiveness check or as part of traceroutes (cf. Section 5.4). Given these differences between the two scanner types, we recommend to use different strategies to study them: for late scanners, DNS-aware IPv6 telescopes are beneficial because they allow researchers to identify distributed scan sources (as described in the preceding paragraph). As early scanners do not reuse DNS responses across scans and also focus more on port scans, studying them with (IPv6-capable) honeypots is possibly a better choice. This allows one to investigate the scanners' layer-7 payloads as well, and, for those scanners that are layer-3 agnostic, honeypots can also be used to record their IPv4 source addresses—two aspects that are especially interesting to those studying these scanners with the intention of creating blocklists.

(iv) **How Should I Expose My Domain Names?** We have shown that it can be worthwhile to expose domain names in multiple sources to attract the greatest number of scanners possible. gTLDs, Tranco, the .ch TLD, and CT logs all attract a substantial number of scanners, including several scanners that only scan domains from a single source. For this reason, we recommend using multiple sources to expose domains. Also, if researchers *deliberately* want to also study scanners relying on the IPv6 Hitlist, they can also attract a fair amount of scanning traffic by getting their addresses included in the Hitlist on purpose. (However, we recently observed that—at least in regard to some protocols—Hitlist scanners are predominantly non-malicious [17].)

(v) **Beware of "Researching the Researchers."** In course of our measurements, we also received a substantial amount of traffic from well-known scanning organizations, some of which operate from multiple ASes like in the CAIDA Ark event in Section 5 or the IPinfo prefix detection measurement in Appendix D. Had we not identified these organizations, we could have easily mistaken these as scans from distinct entities. Hence, it is important to check for well-known scanning organizations as these often send a lot of packets that are not representative of actual malicious scanning traffic. Possible ways of doing this are to identify distributed scans that target the same addresses (see our second point above) and to check whether the source addresses of scans can be attributed to scanning organizations, for example through rDNS records or WHOIS entries. Conversely, researchers and organizations *performing* IPv6 scans should do their best to identify themselves—this includes adhering to best practices like setting up web pages and rDNS records explaining the benign nature of the scans, and giving network operators a chance to opt out [10].

7.2 Limitations & Future Work

Other Hitlists. In theory, it is possible for telescope addresses to appear on other IPv6 hitlists apart from the one operated by TUM. However, hitlists are generally tailored towards collecting *responsive* addresses that can be used in scans. Therefore, if the telescope addresses are unresponsive—as they were in our study—we do not expect hitlists maintainers to include them. The TUM IPv6 Hitlist—to the best of our knowledge—is unique in this regard as it not only publishes *responsive* addresses, but also the cumulative input list of *all* IPv6 addresses found. Nevertheless, should one of or addresses

appear on other hitlists, our one-address-per-query policy would allow us to identify scanners that discovered the address through a common third party. In course of our measurement period, we saw no signs of scanners scanning a common address that likely ended up on a hitlist.

Passive DNS Data. In addition to hitlists, it is also possible for the domains to appear in *passive DNS datasets* if resolver operators collect the queries for our exposed domain names [52]. Unlike hitilsts, this is something that we cannot account for as we know neither which DNS resolvers scanners are using nor which of them collect such data.

Scanners Using Generative Algorithms for IPv6 Addresses. In this study, we have focused on IPv6 scanners targeting the addresses handed out by our DNS server. However, scanners do not have to limit their scans to the addresses we expose: they could—in theory—also use Target Generation Algorithms (TGAs) to search for active addresses in the adjacent address space [6, 8, 45, 46, 50, 53]. While such scans are not the main focus of our work, we investigate this possibility in Appendix D.

Which Scanners Are "Malicious"? Even though we see several scanners sending packets to ports with services they could be trying to exploit (e.g., the web and SSH scans described in Section 5.4), we cannot decisively say if any of these scanners are genuinely malicious. Since our method relies on observing traffic at telescopes that scanners cannot interact with, we, like most of the work on IPv6 scanners up until this point [11, 34, 48, 49, 54], cannot fully judge which scanners harbor malign intentions. However, we have shown that a substantial portion of the IPv6 traffic we observe is clearly *non-malicious* as it originates from well-known scanning organizations. Future work should therefore pivot from analyzing IPv6 scan traffic in general and delve deeper into what IPv6 scan operators do when they manage to connect to a host. As an important first step, we have shown that most domain-based IPv6 scanners target web protocols, often without caring if they connect over IPv6 or IPv4 (cf. Section 5.5). Furthermore, our analysis of layer-3-agnostic scanners revealed that some scanners are crawling for resources related to vulnerabilities (cf. Table 4, Section 5.5). To accurately study if they also exploit these, one could set up a honeypot that can emulate the services the scanners are targeting. This would allow one to observe how scanners interact with these services, and to assess if these interactions are actually malicious.

8 Conclusion

Domain names are an important source for IPv6 scanning campaigns. Through our DNS-aware IPv6 telescope, we are able to uncover a new dimension to IPv6 scans on the Internet—domain-based IPv6 scanners can be separated into two categories: those that scan domains immediately, and those that collect addresses and scan later. As our findings showed, depending on which of these two strategies scanners use, their behavior differs considerably—both in terms of how they conduct their scans and what they are scanning for. We also addressed challenges affecting previous research on IPv6 scanning: by investigating shared destination addresses, we were able to identify scanning campaigns that are spread out across multiple IP addresses—both within and across ASes. We further showed that the IPv6 Hitlist can have a considerable impact when measuring IPv6 scanners—even with non-responsive Telescopes—which needs to be taken into account by future studies.

Researchers operating IPv6 telescopes or honeypots can benefit from our work: they can choose suitable domain exposure methods based on our findings, use our new DNS-aware telescope to uncover distributed scans, and tailor their future investigations towards either the (possibly dual-stack) service-oriented early scanners or the reachability-oriented late scanners.

Going forward, as we find the distinction between different address collection strategies among IPv6 scanners to be of key importance for future studies, we release the source code of our DNS-aware telescope setup as an artifact—thereby allowing other researchers trying to develop IPv6 telescopes to build on our work.

References

[1] Aniket Anand, Michalis Kallitsis, Jackson Sippe, and Alberto Dainotti. 2023. Aggressive Internet-Wide Scanners: Network Impact and Longitudinal Characterization. In *Proc. ACM Int. Conference on emerging Networking EXperiments and Technologies (CoNEXT)*. https://doi.org/10.1145/3624354.3630583 (Cited on pp. 2 and 5)

[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* 48, 2 (May 2018), 2–9. https://doi.org/10.1145/3213232.3213234 (Cited on p. 2)

[3] CAIDA. 2025. *Archipelago (Ark) Measurement Infrastructure.* Retrieved 2025-11-11 from https://www.caida.org/projects/ark/ (Cited on pp. 5 and 10)

[4] Cloudflare. 2025. *WARP.* Retrieved 2025-12-11 from https://developers.cloudflare.com/cloudflare-one/team-and-resources/devices/warp/ (Cited on p. 11)

[5] M. Patrick Collins, Alefiya Hussain, and Stephen Schwab. 2023. Identifying and Differentiating Acknowledged Scanners in Network Traffic. In *Proc. IEEE European Symposium on Security and Privacy (EuroS&P)*. https://doi.org/10.1109/EuroSPW59978.2023.00069 (Cited on pp. 2, 5, and 10)

[6] Tianyu Cui, Gaopeng Gou, Gang Xiong, Chang Liu, Peipei Fu, and Zhen Li. 2021. 6GAN: IPv6 Multi-Pattern Target Generation via Generative Adversarial Nets with Reinforcement Learning. In *Proc. IEEE Int. Conference on Computer Communications (INFOCOM)*. 1–10. https://doi.org/10.1109/INFOCOM42981.2021.9488912 (Cited on p. 18)

[7] Jakub Czyz, Kyle Lady, Sam G. Miller, Michael Bailey, Michael Kallitsis, and Manish Karir. 2013. Understanding IPv6 Internet Background Radiation. In *Proc. ACM Internet Measurement Conference (IMC)*. https://doi.org/10.1145/2504730.2504732 (Cited on p. 2)

[8] Markus Dahlmanns, Felix Heidenreich, Johannes Lohmöller, Jan Pennekamp, Klaus Wehrle, and Martin Henze. 2024. Unconsidered Installations: Discovering IoT Deployments in the IPv6 Internet. In *Proc. IEEE Network Operations and Management Symposium (NOMS)*. 1–8. https://doi.org/10.1109/NOMS59830.2024.10574963 (Cited on p. 18)

[9] Zakir Durumeric, Michael Bailey, and J. Alex Halderman. 2014. An Internet-Wide View of Internet-Wide Scanning. In *Proc. USENIX Security Symposium*. (Cited on pp. 2 and 5)

[10] Zakir Durumeric, Eric Wustrow, and J. Alex Halderman. 2013. ZMap: Fast Internet-wide Scanning and Its Security Applications. In *Proc. USENIX Security Symposium*. 605–620. https://www.usenix.org/conference/usenixsecurity13/technical-sessions/paper/durumeric (Cited on pp. 2 and 17)

[11] Isabell Egloff, Raphael Hiesgen, Maynard Koch, Thomas C. Schmidt, and Matthias Wählisch. 2025. A Detailed Measurement View on IPv6 Scanners and Their Adaption to BGP Signals. *Proc. ACM Netw.* 3, CoNEXT3 (2025). https://doi.org/10.1145/3749215 (Cited on pp. 2, 5, 6, 10, 11, and 18)

[12] Dario Ferrero, Enrico Bassetti, Harm Griffioen, and Harm Smaragdakis. 2025. Have you SYN What I See? Analyzing TCP SYN Payloads in the Wild. In *Proc. ACM Internet Measurement Conference (IMC)*. https://doi.org/10.1145/3730567.3764498 (Cited on pp. 2 and 5)

[13] 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 *Proc. Passive and Active Measurement (PAM)*. https://doi.org/10.1007_978-3-319-76481-8_13 (Cited on p. 4)

[14] Oliver Gasser, Quirin Scheitle, Pawel Foremski, Qasim Lone, Maciej Korczyński, Stephen D. Strowes, Luuk Hendriks, and Georg Carle. 2018. Clusters in the Expanse: Understanding and Unbiasing IPv6 Hitlists. In *Proc. ACM Internet Measurement Conference (IMC)*. https://doi.org/10.1145/3278532.3278564 (Cited on pp. 2, 3, 5, 6, and 8)

[15] Fernando Gont and Tim Chown. 2016. Network Reconnaissance in IPv6 Networks. RFC 7707. https://doi.org/10.17487/RFC7707 (Cited on p. 2)

[16] Harm Griffioen, Georgios Koursiounis, Georgios Smaragdakis, and Christian Doerr. 2024. Have you SYN me? Characterizing Ten Years of Internet Scanning. In *Proc. ACM Internet Measurement Conference (IMC)*. https://doi.org/10.1145/3646547.3688409 (Cited on pp. 2 and 5)

[17] Tiago Heinrich, Sebastian Kappes, and Rafael R. Obelheiro. 2026. Lessons from an (Unsuccessful?) IPv6 Amplification Honeypot Deployment. In *Proc. Workshop on Negative Results in Network Measurements (NetNeg)*. https://doi.org/10.1145/3789240.3828833 (Cited on p. 17)

[18] Raphael Hiesgen, Marcin Nawrocki, Marinho Barcellos, Daniel Kopp, Oliver Hohlfeld, Echo Chan, Roland Dobbins, Christian Doerr, Christian Rossow, Daniel R. Thomas, Mattijs Jonker, Ricky Mok, Xiapu Luo, John Kristoff, Thomas C. Schmidt, Matthias Wählisch, and kc claffy. 2024. The Age of DDoScovery: An Empirical Comparison of Industry and Academic DDoS Assessments. In *Proc. ACM Internet Measurement Conference (IMC)*. https://doi.org/10.1145/3646547.3688451 (Cited on p. 5)

[19] Florian Holzbauer, Markus Maier, and Johanna Ullrich. 2024. Destination Reachable: What ICMPv6 Error Messages Reveal About Their Sources. In *Proc. ACM Internet Measurement Conference (IMC)*. https://doi.org/10.1145/3646547.3688420 (Cited on pp. 2, 9, and 11)

[20] IPinfo. 2025. *IP ASN Dataset.* Retrieved 2025-11-11 from https://ipinfo.io/data/ip-asn (Cited on p. 14)

[21] IPinfo. 2025. *Privacy Detection Dataset.* Retrieved 2025-11-11 from https://ipinfo.io/data/proxy-vpn-detection (Cited on p. 11)

[22] Michael Klopsch, Constantin Sander, Klaus Wehrle, and Markus Dahlmanns. 2025. Time To Scan: Digging into NTP-based IPv6 Scanning. In *Proc. ACM Internet Measurement Conference (IMC).* https://doi.org/10.1145/3730567.3764502 (Cited on p. 6)

[23] Brian Kondracki, Johnny So, and Nick Nikiforakis. 2022. Uninvited Guests: Analyzing the Identity and Behavior of Certificate Transparency Bots. In *Proc. USENIX Security Symposium.* https://www.usenix.org/conference/usenixsecurity22/presentation/kondracki (Cited on p. 6)

[24] Ben Laurie. 2014. Certificate transparency. *Commun. ACM* 57, 10 (Sept. 2014), 40–46. https://doi.org/10.1145/2659897 (Cited on p. 4)

[25] Alexander Lex, Nils Gehlenborg, Hendrik Strobelt, Romain Vuillemot, and Hanspeter Pfister. 2014. UpSet: Visualization of Intersecting Sets. *IEEE Transactions on Visualization and Computer Graphics* (2014). https://doi.org/10.1109/TVCG.2014.2346248 (Cited on p. 12)

[26] Xigao Li, Babak Amin Azad, Amir Rahmati, and Nick Nikiforakis. 2021. Good Bot, Bad Bot: Characterizing Automated Browsing Activity. In *Proc. IEEE Symposium on Security and Privacy (S&P).* https://doi.org/10.1109/SP40001.2021.00079 (Cited on pp. 2 and 6)

[27] Mark Lindsey. 2009. *Never use DNS TTL of zero (0).* Retrieved 2025-11-12 from https://web.archive.org/web/20240215030243/https://mark.lindsey.name/2009/03/09/never-use-dns-ttl-of-zero-0/ (Cited on p. 8)

[28] Alexander Männel, Jonas Mücke, K. C. Claffy, Max Gao, Ricky K. P. Mok, Marcin Nawrocki, Thomas C. Schmidt, and Matthias Wählisch. 2025. Lessons Learned from Operating a Large Network Telescope. In *Proc. ACM SIGCOMM.* https://doi.org/10.1145/3718958.3754347 (Cited on pp. 2 and 5)

[29] David Moore, Colleen Shannon, Geoffrey M Voelker, and Stefan Savage. 2004. *Network Telescopes: Technical Report.* Technical Report. Technical Report CS2004-0795, CSE Department, UCSD. (Cited on pp. 2 and 5)

[30] Cristian Munteanu, Said Jawad Saidi, Oliver Gasser, Georgios Smaragdakis, and Anja Feldmann. 2023. Fifteen Months in the Life of a Honeyfarm. In *Proc. ACM Internet Measurement Conference (IMC).* 282–296. https://doi.org/10.1145/3618257.3624826 (Cited on p. 2)

[31] Ruoming Pang, Vinod Yegneswaran, Paul Barford, Vern Paxson, and Larry Peterson. 2004. Characteristics of Internet Background Radiation. In *Proc. ACM Internet Measurement Conference (IMC).* https://doi.org/10.1145/1028788.1028794 (Cited on pp. 2 and 5)

[32] Stijn Pletinckx, Thanh-Dat Nguyen, Tobias Fiebig, Christopher Kruegel, and Giovanni Vigna. 2023. Certifiably Vulnerable: Using Certificate Transparency Logs for Target Reconnaissance. In *Proc. IEEE European Symposium on Security and Privacy (EuroS&P).* https://doi.org/10.1109/EuroSP57164.2023.00053 (Cited on pp. 2, 4, and 6)

[33] Victor Le Pochat, Tom van Goethem, Samaneh Tajalizadehkhoob, Maciej Korczynski, and Wouter Joosen. 2019. Tranco: A Research-Oriented Top Sites Ranking Hardened Against Manipulation. In *Proc. Network and Distributed System Security Symposium (NDSS).* (Cited on p. 4)

[34] Philipp Richter, Oliver Gasser, and Arthur Berger. 2022. Illuminating Large-Scale IPv6 Scanning in the Internet. In *Proc. ACM Internet Measurement Conference (IMC).* https://doi.org/10.1145/3517745.3561452 (Cited on pp. 6, 10, and 18)

[35] RIPE. 2025. *RIPE Atlas.* Retrieved 2025-11-11 from https://atlas.ripe.net/ (Cited on p. 5)

[36] Kimberly Ruth, Deepak Kumar, Brandon Wang, Luke Valenta, and Zakir Durumeric. 2022. Toppling Top Lists: Evaluating the Accuracy of Popular Website Lists. In *Proc. ACM Internet Measurement Conference (IMC).* https://doi.org/10.1145/3517745.3561444 (Cited on p. 4)

[37] Erik Rye and Dave Levin. 2023. IPv6 Hitlists at Scale: Be Careful What You Wish For. In *Proc. ACM SIGCOMM.* https://doi.org/10.1145/3603269.3604829 (Cited on pp. 2 and 4)

[38] Erik C. Rye and Robert Beverly. 2020. Discovering the IPv6 Network Periphery. In *Proc. Passive and Active Measurement (PAM)*, Anna Sperotto, Alberto Dainotti, and Burkhard Stiller (Eds.). 3–18. (Cited on p. 2)

[39] Quirin Scheitle, Oliver Gasser, Theodor Nolte, Johanna Amann, Lexi Brent, Georg Carle, Ralph Holz, Thomas C. Schmidt, and Matthias Wählisch. 2018. The Rise of Certificate Transparency and Its Implications on the Internet Ecosystem. In *Proc. ACM Internet Measurement Conference (IMC).* https://doi.org/10.1145/3278532.3278562 (Cited on p. 6)

[40] Quirin Scheitle, Oliver Hohlfeld, Julien Gamba, Jonas Jelten, Torsten Zimmermann, Stephen D. Strowes, and Narseo Vallina-Rodriguez. 2018. A Long Way to the Top: Significance, Structure, and Stability of Internet Top Lists. In *Proc. ACM Internet Measurement Conference (IMC).* https://doi.org/10.1145/3278532.3278574 (Cited on p. 4)

[41] David Schinazi and Tommy Pauly. 2017. Happy Eyeballs Version 2: Better Connectivity Using Concurrency. RFC 8305. https://doi.org/10.17487/RFC8305 (Cited on p. 7)

[42] Raffaele Sommese, Gautam Akiwate, Antonia Affinito, Moritz Muller, Mattijs Jonker, and kc claffy. 2024. DarkDNS: Revisiting the Value of Rapid Zone Update. In *Proc. ACM Internet Measurement Conference (IMC).* https://doi.org/

10.1145/3646547.3689021 (Cited on p. 4)

[43] Raffaele Sommese, KC Claffy, Roland van Rijswijk-Deij, Arnab Chattopadhyay, Alberto Dainotti, Anna Sperotto, and Mattijs Jonker. 2022. Investigating the impact of DDoS attacks on DNS infrastructure. In *Proc. ACM Internet Measurement Conference (IMC)*. https://doi.org/10.1145/3517745.3561458 (Cited on p. 5)

[44] Raffaele Sommese, Roland van Rijswijk-Deij, and Mattijs Jonker. 2024. This Is a Local Domain: On Amassing Country-Code Top-Level Domains from Public Data. *ACM SIGCOMM Computer Communication Review* (2024). (Cited on p. 4)

[45] Lion Steger, Liming Kuang, Johannes Zirngibl, Georg Carle, and Oliver Gasser. 2023. Target Acquired? Evaluating Target Generation Algorithms for IPv6. In *Proc. Network Traffic Measurement and Analysis Conference (TMA)*. (Cited on pp. 18 and 24)

[46] Lion Steger, Liming Kuang, Johannes Zirngibl, Georg Carle, and Oliver Gasser. 2026. Still on Target? An Evaluation of IPv6 Target Generation Algorithms. *IEEE Transactions on Network and Service Management* (2026). https://doi.org/10.1109/TNSM.2026.3705935 (Cited on pp. 18 and 24)

[47] Stephen D Strowes, Emile Aben, René Wilhelm, Florian Obser, Riccardo Stagni, and Agustin Formoso. 2020. Debogonising 2a10::/12: Analysis of One Week's Visibility of a New/12.. In *Proc. Network Traffic Measurement and Analysis Conference (TMA)*. (Cited on p. 2)

[48] Hammas Bin Tanveer, Echo Chan, Ricky K. P. Mok, Sebastian Kappes, Philipp Richter, Oliver Gasser, John Ronan, Arthur Berger, and kc Claffy. 2025. Unveiling IPv6 Scanning Dynamics: A Longitudinal Study Using Large Scale Proactive and Passive IPv6 Telescopes. *Proc. ACM Netw.* 3, CoNEXT3 (2025). https://doi.org/10.1145/3749221 (Cited on pp. 2, 5, 6, 10, and 18)

[49] Hammas Bin Tanveer, Rachee Singh, Paul Pearce, and Rishab Nithyanand. 2023. Glowing in the Dark: Uncovering IPv6 Address Discovery and Scanning Strategies in the Wild. In *Proc. USENIX Security Symposium*. (Cited on pp. 2, 5, 6, and 18)

[50] Grant Williams and Paul Pearce. 2024. Seeds of Scanning: Exploring the Effects of Datasets, Methods, and Metrics on IPv6 Internet Scanning. In *Proc. ACM Internet Measurement Conference (IMC)*. https://doi.org/10.1145/3646547.3688449 (Cited on pp. 18 and 24)

[51] Eric Wustrow, Manish Karir, Michael Bailey, Farnam Jahanian, and Geoff Huston. 2010. Internet Background Radiation Revisited. In *Proc. ACM Internet Measurement Conference (IMC)*. https://doi.org/10.1145/1879141.1879149 (Cited on pp. 2 and 5)

[52] Yunpeng Xing, Chaoyi Lu, Baojun Liu, Haixin Duan, Junzhe Sun, and Zhou Li. 2024. Yesterday Once More: Global Measurement of Internet Traffic Shadowing Behaviors. In *Proc. ACM Internet Measurement Conference (IMC)*. 230–240. https://doi.org/10.1145/3646547.3689023 (Cited on pp. 2 and 18)

[53] Tao Yang, Zhiping Cai, Bingnan Hou, and Tongqing Zhou. 2022. 6Forest: An Ensemble Learning-based Approach to Target Generation for Internet-wide IPv6 Scanning. In *Proc. IEEE Int. Conference on Computer Communications (INFOCOM)*. 1679–1688. https://doi.org/10.1109/INFOCOM48880.2022.9796925 (Cited on p. 18)

[54] Liang Zhao, Satoru Kobayashi, and Kensuke Fukuda. 2024. Exploring the Discovery Process of Fresh IPv6 Prefixes: An Analysis of Scanning Behavior in Darknet and Honeynet. In *Proc. Passive and Active Measurement (PAM)*. (Cited on pp. 2, 5, 6, and 18)

[55] Johannes Zirngibl, Lion Steger, Patrick Sattler, Oliver Gasser, and Georg Carle. 2022. Rusty Clusters? Dusting an IPv6 Research Foundation. In *Proc. ACM Internet Measurement Conference (IMC)*. https://doi.org/10.1145/3517745.3561440 (Cited on pp. 2, 3, 5, 6, and 8)

A Ethics

The traffic collected by our telescope does not originate from human users but from automated scanners. We therefore do not collect any personal data or traffic from actual Internet users. The nameserver and telescope are dedicated setups that are not used for any other services. Hence, they do not attract traffic from real users. Similarly, our exposed domains are newly registered and not used for any other services.

Our work mostly relies on passive data collection in a telescope, but also uses active scans to identify acknowledged scanners. We conduct rDNS lookups for all IPv6 addresses contacting our telescope and check whether scanners host a website. These lookups are done once per scanner IP address. Thus, the impact in terms of traffic is minimal.

When placing domains on the Trance top list, we took precautions to minimize their influence on research studies that rely on Tranco. As it is common for researchers to scan subsets of higher-ranked domains, we make sure that our honey domains do not make up more than 0.02% of the top 10^n domains (for all $n \in \mathbb{N}$) at any point in time. For example, we only insert at most two of our domains into the top 10k. This means that anyone scanning the top 10k will only scan two of our domains—consequently, we would not have a *statistically significant* impact on their research findings. Depending on the research question, the volatility of top lists (i.e., the churn of domains within the top list) and the unavailability of domains within them will have a much larger impact on experiments. Furthermore, we always place our domains in the lower-ranked half of the respective top 10^n domains, and we do not place any domains in the top 9k. As a consequence, none of our domains are ever included in the top 0.9% (i.e., the top 9K) of the Tranco Top 1M. We consider this a one-off experiment and do not plan to include experiment domain names in Tranco long-term.

Similarly, we only put a small set of addresses on the IPv6 Hitlist, limiting the impact on studies using it as input for IPv6 scans.

We discussed our approach with our Institutional Review Board (IRB) and they approved this study.

B Open Science

To aid other researchers studying IPv6 scanners, we release all traffic captured by our telescope and our preprocessed traffic data with annotations (e.g., domain sources, Δt values and tags for known scanning organizations) as an artifact. Also, as we want to enable others to deploy their own DNS-aware IPv6 telescopes as well, we provide the source code and setup instructions for our telescope-aware DNS server that allows correlating resolution times with telescope traffic. Both the source code and the data mentioned are available at `https://doi.org/10.17617/3.JS9HCI`.

C Distributed Scanning Group in AS14061 (DigitalOcean)

Figure 9 visualizes the temporal development of scanners in AS14061 (DigitalOcean) discussed in Section 5.1. Different groups of addresses perform scans every Monday for 3-5 weeks. Afterwards, the set of addresses changes but the behavior remains the same.

D Do Domain-Based IPv6 Scanners Target Unexposed Addresses?

As IPv6 scanners are searching for responsive scan targets in the IPv6 address space, their scans do not necessarily have to be confined to the addresses which we exposed via our DNS server: they might also probe neighboring addresses to uncover other responsive hosts in the same network. We therefore analyze which of the domain-based IPv6 scanners in our telescope also target addresses other than the ones associated with a honey domain names.

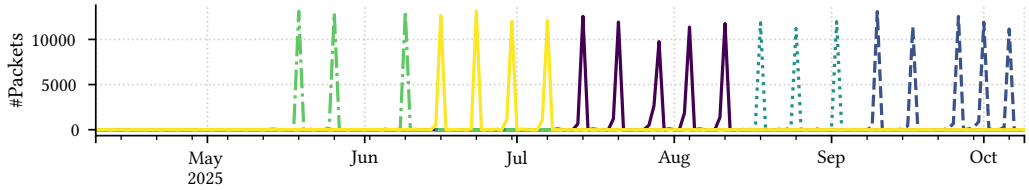


Fig. 9. **Daily scan packets of the five groups of addresses in AS14061 that share destination addresses**—Each color denotes a different group. A single group performs scans every Monday for 3–5 weeks before it stops and another group takes over.

For this, we take the set of all source IP addresses of the scanners that targeted an IP address exposed through our DNS server and check if any of them also sent packets to an address that was not exposed via DNS. We find that of the 25k addresses in 592 ASes targeting exposed addresses in our telescope, 178 addresses in 41 ASes also send traffic to non-exposed addresses in the telescope. Based on the rDNS records, 169 (95%) of these addresses are associated with IPinfo. We were able to confirm that these addresses are indeed part of IPinfo's IPv6 aliased prefix detection measurements. The remaining 9 non-IPinfo addresses were located in three ASes and only sent 34 packets in total.

While we could not find many source addresses scanning both exposed and unexposed telescope addresses, it is possible that our domain names still attracted scanners to our telescope that do not target the exposed addresses themselves. As the goal of our method is to study scanners targeting exposed addresses, identifying such scanners is beyond the scope of this paper; however, we will try to provide a best-effort analysis of this in remainder of this section.

To investigate if our domain experiments attracted traffic from scanners predominantly targeting unexposed addresses, we compare the traffic directed at the unexposed addresses in the upper half of our /36—which contains the addresses that our domains resolve to—to the lower, unused half of the telescope where we do not expose any addresses.[5]

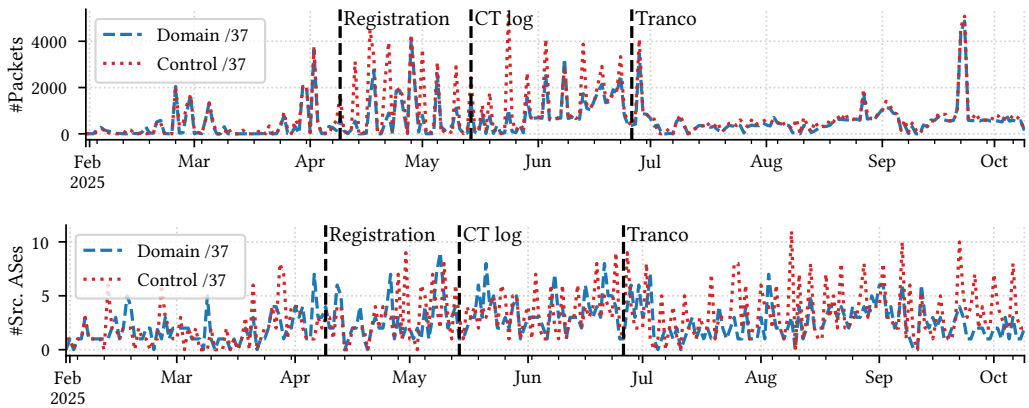


Fig. 10. Daily traffic and source ASes in the upper /37 of our telescope of where we expose addresses vs. the lower /37 where we do not.

In total, our telescope received 1M packets to unexposed addresses coming from 442 addresses in 55 ASes. Figure 10 shows the daily number of packets and source ASes targeting unexposed addresses for both halves of the telescope. We see no clear indication that the /37 where we expose

[5]For this analysis, we exclude the IPinfo addresses that we identified earlier.

our addresses attracts traffic from more ASes than the other half. In fact, the unexposed addresses in the upper half of our telescope saw both less traffic (815K vs. 195K packets), and less source ASes (45 vs. 48 ASes) than those in the lower half. However, the domain /37 received traffic to unused addresses from 7 ASes that did not contact the control /37. Moreover, we observe an intersection of 38 ASes in both halves of the telescope, 13 of which only contacted the telescope *after* we started exposing addresses via domain names. The average number of daily packets to unexposed addresses also increases after the begin of our experiments. As our current method does not allow us to make any inferences about whether our domain names might have attracted these scans, we leave this question to future work.

Key Takeaways: *We find several source addresses that scan both exposed and unexposed address in our telescope. Most of these belong to IPinfo—a known scanning organization. For the source addresses that* only *target unexposed addresses, we see no clear indication that they were more likely to target the half of the telescope where we exposed addresses.*

E Additional Analysis of the Hitlist Scanners

As we expose addresses in multiple datasets published by the Hitlist, we can also study the combination of datasets used by the source ASes we observe. Figure 11 shows the intersections between the different datasets and the number of scan sources they attract. In this plot we combined the eight addresses exposed in each dataset into one category each (cf. Section 4.4). The only exception is `icmp_rand_1`, the first of two random addresses published in the list for addresses responsive to ICMPv6 probes, which we choose not to combine with the other seven addresses in this dataset as it attracted exceptionally many source ASes. This is likely due to a distributed scanner targeting only a subset of the addresses in the *icmp* list, e.g., by filtering to only include a number of target address per subnet.

The most-used dataset in terms of attracted source ASes is the Hitlist `input` dataset, which currently holds more than four billion addresses and is the most extensive dataset. As this is a cumulative list of collected addresses going back to 2018, most addresses are historic and currently unreachable (for comparison the most recent list of responsive addresses contains only 20.6 M addresses). This is followed by the `icmp_rand_1` address from the `icmp` dataset. Both datasets require a registration at the service. The `publicly` available dataset of responsive addresses was used by scanners from 21 source ASes. All remaining ASes use a combination of datasets, including the combination of all available datasets, except for the `udp53` dataset, which two ASes use exclusively. Usage of the latter indicates dedicated DNS measurements.

Furthermore, we find a small percentage of probes (1.36 %, 8006 probes) that do not target the exact published addresses, but other addresses within the same /64 prefixes which we allocated for the experiment addresses. 2290 probes (28.6 %) originated from the TUM Hitlist aliased prefix detection mechanism and were caused by a configuration error on behalf of the maintainers. The remaining probes originated from 16 different ASes, with the top three ASes accounting for 92 % of probes. 97 % of probes use ICMPv6 echo requests with 2.8 % using TCP and eight probes using UDP. We find that some of the probed addresses follow common patterns, where the values of some address nibbles are randomized whereas others stay fixed. This indicates that scanners might use heuristic scanning methods to discover unknown active addresses by identifying patterns in known addresses. This is a common technique employed by TGAs, which are also used for Internet measurements and predominantly send ICMPv6 echo requests for probing [45, 46, 50].

Lastly, by manually inspecting the source ASes and addresses as well as their rDNS records, we were able to map 78.76 % of all probes (excluding those which do not target the exact published addresses) to known scanning organizations. We list the organizations we were able to identify in Table 6.

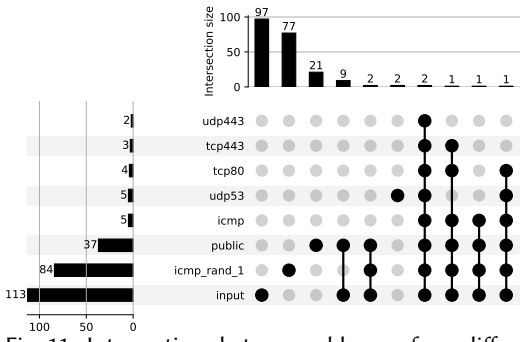


Fig. 11. Intersections between addresses from different IPv6 Hitlist datasets used by different source ASes.

Table 6. Organizations scanning the addresses exposed on the IPv6 Hitlist.

Organization	#Src. ASes	#Probes	#Dst. Ports	Used datasets
BinaryEdge	1	3489	245	public
CAIDA	8	8	0	icmp_rand_1
Censys	1	6	1	public
FH Münster	1	5046	38	public
Georgia Tech	1	50	5	public
Driftnet	1	237 101	8408	public
IPinfo	2	190	0	input
RIPE Atlas	4	172	1	input, icmp_rand_1
Shadowserver	1	205 917	114	public, input
University of Twente	1	3087	1	icmp_rand_1
Unknown	204	124 840	3378	public, input, udp53 icmp, icmp_rand_1 tcp443, tcp80, udp443
Total	213	579 907	8863	public, input, udp53 icmp, icmp_rand_1 tcp443, tcp80, udp443

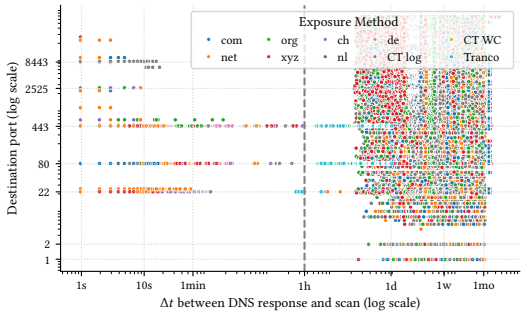


Fig. 12. **Destination port of TCP/UDP packets sent to the telescope vis-à-vis their** Δt—Each dot represents a packet. Most of the late port scans originate from single AS that targets 43.9K different ports (AS211298/Driftnet).

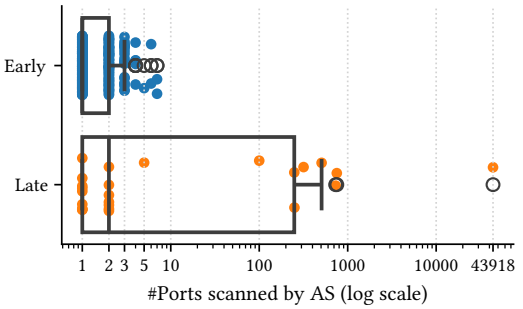


Fig. 13. **Strip and box plot visualizing the number of ports scanned by early vs. late ASes that send TCP/UDP traffic**—Each colored dot represents an AS. ASes target a greater range of ports in late scans.

F Differences in Port Scan Behavior between Early and Late Scanners

To give a better overview of which ports early and late scans target, Figure 12 shows the destination ports of all TCP/UDP packets sent to the telescope in relation to their Δt. The left side of the plot contains the early scans: overall, early scans target only 18 unique destination ports, with ports 80 and 443 being the most popular (cf. Section 5.4). Moving further to the right, the first late scans start to appear: a prominent example are the BYU scans targeting ports 80 and 443 of Tranco domains (cf. Section 5.1), which occur between 89min and 24.5h after the target address was handed out by the DNS server. The rightmost portion of the plot is dominated by the scans of Driftnet, our most prominent source of traffic (cf. Section 5.1). Driftnet exclusively performs late scans, which target 43.9K different ports and have a Δt range from 6.2h to 32 days.

Although only 25 late scanners sent TCP/UDP packets to our telescope, 8 (32%) of them target more unique ports on their own than all early scanners combined. Figure 13 further illustrates this discrepancy: late TCP/UDP scanners target more ports than their early counterparts. In addition to Driftnet with its 43.9K destination ports, 7 other ASes perform late scans of more than 100 ports. Opposed to these are the early scans, where the largest number of ports scanned by a single AS is 7.

Received 12 December 2025; revised 29 May 2026; accepted 22 June 2026