Unconsidered Installations: Discovering IoT Deployments in the IPv6 Internet

Markus Dahlmanns*, Felix Heidenreich*, Johannes Lohmöller*,
Jan Pennekamp*, Klaus Wehrle*, and Martin Henze§,‡
*Communication and Distributed Systems, RWTH Aachen University, Germany
§Security and Privacy in Industrial Cooperation, RWTH Aachen University, Germany
‡Cyber Analysis & Defense, Fraunhofer FKIE, Germany
{dahlmanns, heidenreich, lohmoeller, pennekamp, wehrle}@comsys.rwth-aachen.de · henze@cs.rwth-aachen.de

***Abstract*—Internet-wide studies provide extremely valuable insight into how operators manage their Internet of Things (IoT) deployments in reality and often reveal grievances, e.g., significant security issues. However, while IoT devices often use IPv6, past studies resorted to comprehensively scan the IPv4 address space. To fully understand how the IoT and all its services and devices is operated, including IPv6-reachable deployments is inevitable—although scanning the entire IPv6 address space is infeasible.**

In this paper, we close this gap and examine how to discover IPv6-reachable IoT deployments. Using three sources of active IPv6 addresses and eleven address generators, we discovered 6658 IoT deployments. We derive that the available address sources are a good starting point for finding IoT deployments. Additionally, we show that using two address generators is sufficient to cover most found deployments. Assessing the security of the deployments, we surprisingly find similar issues as in the IPv4 Internet, although IPv6 deployments might be newer and generally more up-to-date: Only 39 % of deployments have access control in place and only 6.2 % make use of TLS inviting attackers, e.g., to eavesdrop sensitive data.

***Index Terms*—Internet of Things, Internet measurements, IPv6**

I. Introduction

The Internet of Things (IoT) with all its derivatives, e.g., the Industrial IoT (IIoT) or smart homes, offers various benefits to users and society alike [1], [2]. To realize these benefits, corresponding deployments often interact with the physical world and have to process and communicate highly sensitive data [3], [4]. Hence, operators must run their deployments with care, e.g., regarding security measures like access control. For communication, the growing number of IoT deployments benefits from the significant size of the IPv6 address space since IPv4 addresses are rare, depleted, and thus numerically insufficient for all Internet-connected devices [5], [6].

Internet-wide studies allow an understanding of how operators manage real and existing Internet-facing deployments at scale, e.g., [7]–[9]. Thus, their results are indispensable for the standardization of network protocols and the derivation of requirements for intelligent network management tools. While Internet-wide studies in the complete IPv4 address space (2^{32} addresses) finish in a few minutes [10], they are not feasible for all 2^{128} IPv6 addresses as measurements would require 600 trillion years at the same speed (10 Gbit/s). Thus, fundamentally new approaches are needed to understand how operators manage their IoT services in the wild.

State-of-the-art Internet measurement methods resort to scanning only parts of the IPv6 address space by relying on (i) *hitlists* [11], [12] of active IPv6 addresses from different sources, e.g., DNS or email, and (ii) numerous *generators*, e.g., [13]–[15], that take seeds, such as hitlists, to produce further IPv6 addresses which might be in use and thus are valuable to scan. While these approaches tend to work well finding Web services [12], [16], their practicability for IoT services is not well researched. Past scans only report a comparably low number of IPv6-reachable IoT(-backend) services [17], [18] while relying on a single hitlist. Thus, it is unknown how well generators combined with which seeds can increase the number of discoverable IoT deployments (as for Web services). However, answering this question is key to properly understand whether operators misconfigure IPv6-reachable IoT deployments similarly to deployments in the IPv4 address space [7], [19], [20], or whether these deployments profit from their (potentially) more recent installation.

In this paper, we thus address the research gap of understanding the findability, prevalence, and configuration of IoT deployments in the IPv6 address space. More specifically, we combine eleven open-sourced generators and three seedlists to discover IoT services in the IPv6 address space that use common IoT protocols, i.e., AMQP, MQTT, OPC UA, and CoAP, determine which combination of address sources works best to find IoT deployments, and exemplarily assess their security configuration in comparison to the IPv4 address space.

Contributions: Our main contributions are as follows.

- We adapt eleven state-of-the-art address generators for IoT-focused scanning on three seedlists (from related work, DNS, and data from IPv4 scans) and find 6658 IPv6-reachable IoT deployments.
- Tracking the origin of the IPv6 addresses, we derive that using all seedlists for scanning is beneficial. Still, two generators suffice to find 95 % of all identified IoT deployments.
- We show that the security configuration IPv6-reachable IoT deployments does not differ from their IPv4 counterparts, i.e., they show the same problems such as missing communication security (94 %) and disabled access control (48 %).
- We open-source our used tools to scan and track IPv6 addresses to support future research [21].

		Name	Year	Technique
Lists		Song et al. [29]	2020	—
		TUM [11], [12], [25]	2016 / 2022	—
Scanlist Generation	Passive	Ullrich et al. [30]	2015	Partial Pattern Discovery
		Entropy/IP [31]	2016	Bayesian Network (Entropy, Random)
		6Gen [32]	2017	High Density Region Fill Up
		EIP-Generator [11]	2018	Bayesian Network (Entropy, Complete)
		IEDC [33]	2020	High Density Region Fill Up
		6GCVAE [34]	2020	Variational Autoencoder
		6VecLM [35]	2020	Language Model
		6Graph [13]	2021	High Density Region Fill Up (Graph)
		6GAN [36]	2021	Generative Adversarial Network
		6Forest [37]	2022	Isolation Forest
	Active	6Tree [38]	2019	Diverse Hierarchical Clustering (DHC)
		6Hit [39]	2021	Reinforcement Learning
		DET [14]	2022	DHC (incl. density and hierarchies)
		AddrMiner [40]	2022	Balanced Spatial Pattern Representation
		6Scan [15]	2023	Regional Encoding

TABLE I: Numerous approaches try to ease IPv6 scanning. Approaches in gray are (openly) available.

II. Related Work & Background

Our measurements of IPv6-reachable IoT services are motivated by approaches guiding scanners in the IPv6 Internet to addresses of interest and related work analyzing the operation of IoT deployments in the IPv4 address space.

A. IPv6-wide Scanning

While scans for the entire IPv4 space finish within minutes [10], completely scanning the larger IPv6 address space is infeasible [22]. Thus, as listed in Table I, various approaches attempt to identify IP addresses worth scanning.

Hitlists: Early on, researchers proposed to base IPv6 scanning on resources that contain addresses known to be in use [22]–[24], e.g., based on passive Internet traffic, or DNS, and evaluated how many active IPv6 addresses respective sources can derive [25], [26]. To provide a starting point for IPv6 scanning, researchers from TUM [11] make their addresses available in a hitlist. The public availability already backed numerous works, e.g., analyzing the deployment of QUIC [27] or peripheral devices [28].

Scanlist Generation: By analyzing hitlists, Gasser et al. revealed that used IP addresses are clustered [11]. Consequently, generation approaches for the IPv6 address space emerged to enable more targeted scanning (cf. Table I). While passive approaches take a *seedlist* to generate *scanlists*, active strategies scan selected IP addresses to incorporate live network information. The approaches differ in the techniques used to identify new targets, leading to varying effectiveness. Recently, the TUM extended their hitlist with IPs generated by 6GAN, 6Graph, 6Tree, and 6VecLM, but already noted reduced effectiveness for finding Web services compared to their original publication [12].

Alias Detection: Previously, single servers were found to answer requests to complete IPv6 subnets, potentially biasing measurements [11]. Thus, researchers leveraged protocol features of SNMPv3 [41], IPv6 fragmentation [42], [43], ICMP rate limiting [44], the TTL from different vantage points [45], unused addresses [46], and multi-protocol behavior [47] to "dealias" IPs used in the Internet core only. For host addresses, researchers either utilize the caching behavior of Path Maximum Transmission Units ("Too Big Trick") [14] or the sparsity of IPv6 addressing schemes by checking for similar responses in the surrounding subnet of a single IP address [11], [31].

Takeaway: *Numerous works target to increase the coverage of IPv6 scans, allowing researchers to use hitlists and address generators in every possible combination. However, until now, the evaluation of their value heavily focused on Web services.*

B. Past Efforts of IoT Scanning

Different Internet scan services, e.g., Censys [48], [49], perform active measurements to collect and share meta-information on reachable (IoT) deployments [50], [51]. Such meta-information allows attackers to find deployments that are insufficiently configured [52]–[56]. Still, these services do not find all Internet-reachable deployments [57].

For several years now, researchers have overcome this issue by collecting their data on Internet-reachable IoT deployments [4], [19], [20], [58] using, e.g., ZMap [59]. For example, these measurements evinced more than 300 000 MQTT brokers [4], [19], [20] and more than 1000 OPC UA deployments [7]. Security analyses of these deployments reveal that many fail to enable access control or end-to-end security [4], [7], [19], [20], thus opening many doors for attackers.

Strikingly, all of these works focus on the IPv4 address space, i.e., their assessment does not cover the larger (and possibly more modern) IPv6 part of the Internet. Notably, large scanning services only list a handful of IoT-related IPv6 deployments [49], [60] which were classified in a previous work [61]. Even initial works searching IoT(-backend) services using a hitlist (but no generators) [17], [18] only find a limited number. It is thus open whether and which address generators enable a broader view of IoT deployments in the IPv6 Internet.

Takeaway: *Internet-wide studies revealed various peculiarities for IoT deployments but focused on IPv4 so far. Which technique works best to find IPv6-reachable IoT deployments remains unclear but is fundamental to gain a full view.*

III. IoT-Focused IPv6 Scanning

We augment prior efforts on IPv6 scanning by performing active measurements to specifically analyze the spread of IoT and IIoT deployments. To ascertain how well specific address generation techniques and address sources perform, we simultaneously track the origin(s) of each scanned IP address.

A. Methodology

To understand which measurement strategy works best to get insight into how operators manage IPv6-reachable IoT deployments, our methodology combines seedlists and generators with IoT scanning and its ethical needs.

1) Protocol Focus: We focus on four IoT-related protocols that implement modern communication paradigms and were subject to recent studies in the IPv4 address space [7], [19], [20]. Specifically, due to their modernity, our scans cover the two Publish-Subscribe protocols AMQP and MQTT, usually

Section III-A (Methodology)							*Section III-B (Validating Responses)*				*Section III-C (Information on Valid Deployments)*				
	Protocol	**IP**	**Port**	**Variant**	**Date (2023)**	**Scanned IPs**	**Hosts**	**Transport**	**(D)TLS Success**	**Valid**	**ASes Distinct**	**ASes +IPv6**	**ASes Types (Valid (%))[1]**	**Cert. CNs Distinct**	**Aliasing**
Backend	AMQP	IPv4	5672†	Standard TLS	06-04	3 696 506 441	5 300 545	312 883	—	143 200	5362 (0.03/Valid)	—	1. Content (74 %)	—	—
			5671‡	TLS	06-03		5 345 896	216 882	128 476	11 530			2. Enterprise (15 %)	3465	
		IPv6	5672†	Standard TLS	06-06	1 849 337 203	813 263	52 150	162 492	28 359			3. NSP (6.4 %)	(0.09/Valid)	
			5671‡	TLS	06-04	1 799 131 401	2 766 434	50 538	—	4664	318 (0.07/Valid)	33	1. Content (89 %)	—	25
	MQTT	IPv4	1883†	Standard TLS	05-28	3 696 506 441	3 822 281	720 817	38 217	18			2. NSP (5.9 %)	126	
			8883‡	TLS	05-27		6 003 536	704 098	41 327	137			3. ISP (2.1 %)	(0.81/Valid)	
		IPv6	1883†	Standard TLS	05-31	1 828 965 301	796 283	86 772	—	359 970	4060 (0.01/Valid)	—	1. ISP (80 %)	—	—
			8883‡	TLS	05-29	1 800 732 625	1 062 446	186 269	135 081	390			2. Content (10 %)	3938	
Device	OPC UA	IPv4	4840†	Standard TLS	06-11	3 696 506 441	4 085 685	17 616	232 014	12 759			3. NSP (7.4 %)	(0.3/Valid)	
			4843‡	TLS	06-10		5 218 969	18 604	—	1675	221 (0.11/Valid)	8	1. Content (75 %)	—	20
		IPv6	4840†	Standard TLS	06-13	1 849 275 639	2 680 314	731	56 188	11			2. NSP (12 %)	197	
			4843‡	TLS	06-11	1 785 421 878	3 561 978	524	6370	245			3. ISP (8.1 %)	(0.77/Valid)	
	CoAP	IPv4	5683†	Standard DTLS	06-18	3 696 506 441	326 694	269 589	—	1740	491 (0.28/Valid)	—	1. ISP (38 %)	869	—
			5684‡	DTLS	06-17		72 841	19 889	6098	2			2. NSP (35 %)	(0.5/Valid)	
		IPv6	5683†	Standard DTLS	06-24	2 219 964 687	0	0	—	0			3. Content (15 %)		
			5684‡	DTLS	06-23	1 762 033 586	55	24	0	0					
									600	91	3 (0.5/Valid)	0	1. Content (83 %)	0	0
									—	266 037			2. NSP (17 %)	(0/Valid)	
									0	0			3. N/A (0 %)		
									—	0	3067 (0.01/Valid)	—	1. NSP (66 %)	—	—
									0	0			2. ISP (32 %)	43	
									—	0			3. Content (1.3 %)	(0.47/Valid)	
									0	0	2 (1/Valid)	0	1. Content (50 %)	—	0
									10	2			2. Education (50 %)	2	
													3. N/A (0 %)	(1/Valid)	

TABLE II: *Left:* Protocols and their variants (†standard / ‡secure port), scan dates, and number of scanned IP addresses. *Center:* Results of our validation process. *Right:* AS, certificate, and aliasing information on valid deployments.

used for IoT-backend infrastructures [19], [20], as well as CoAP and OPC UA as promising representatives for (I)IoT-related protocols implemented by IoT devices. Since we are also interested in potential different (security) operation of deployments, we scan for both the standard and TLS-secured (DTLS in case of CoAP) variant of these protocols (cf. Table II (left)). Our protocol selection significantly influences active generators as well as our scanning since it requires support for the specific transport and application layer protocol via the respective port.

2) Seedlist Sources: To understand which source of IPv6 addresses performs best in this regard, we rely on three primary sources for IPv6 addresses: (i) the input IP addresses for *TUM hitlist* [11], [12], as it is a widely established source for scanning IPv6 addresses, (ii) AAAA records behind domains (with and without subdomain `www`) included in *DNS Zone files* [62], as operators might rely on easy to remember domain names to connect to IoT-related services, and (iii) AAAA records behind domains set as *RDNS entry* of addresses from or included in *certificates* gathered during a *previous IPv4 scan* on the respective port (cf. Table II (left)).

3) Scanlist Generation: Since IPv6 scanlist generation approaches (cf. Section II) promise to find more deployments in the IPv6 address space, we use them to generate further addresses for our scans. However, the generated address sets of different approaches usually only overlap rarely [12] and, until now, which approach performs best to find IoT-related deployments remains unclear. Thus, we use all open-sourced generation approaches to answer this question.

Generator Configuration: The possible configurations of the generators influencing the output are manifold, e.g., the number of input and output addresses or internally used clustering approaches. However, the (comparably) extensive runtime of the approaches prevents evaluating all possible configurations of all generators while relying on up-to-date seedlists and running the IPv6 scan close after an IPv4 scan to uphold its comparability. Instead, to keep the runtime of the address generation feasible, we use the best-performing configuration from the respective publication. Still, the extensive runtime of 6VecLM forces us to further reduce the number of input addresses to 10 000. To additionally save GPU resources we run 6GAN and 6GCVAE on *TUM* and *v4* sources only.

Generator Input: To feed the generators with the selected amount of seed addresses, we randomly sample seedlists whenever they include more IP addresses as required as input. We run multiple instances of passive generators in parallel on different input samples when the generator's runtime permits and the input list has more entries than randomly selected.

Generator Results: We list the number of scanned IP addresses in Table II (left), which vary per scan due to addresses internally scanned by active approaches and approaches not allowing to set the number of generated target addresses.

4) Scanning IPv6 & IPv4: To evaluate whether IoT services run behind the IPv6 addresses from our seedlists and the subsequent generation results, we use `zmapv6` [63] on ports of our curated list (Table II (left)). For our accompanying IPv4 scans, we scan the entire address space relying on `zmap` [59].

Whenever we find IP addresses with a specific port open, we subsequently use `zgrab2` [64] to perform application layer handshakes and retrieve configuration information as well as payload data. To also find deployments running the (D)TLS protocol variant on the standard port, we further retry establishing a (D)TLS connection when the standard application handshake was unsuccessful.

5) Ethical Considerations: Since our measurements affect and concern real, potentially resource-constrained IoT deployments, we must carefully follow established research guidelines [65], best practices for Internet-wide measurements [59], and regulations enacted by our institutions.

Implications for IoT Deployments: First, we ensure that we do not send requests to single IPs too frequently as this might overload IoT deployments. Here, we must consider two spots in our methodology: (i) IP addresses occurring in the output of more than one generator, and (ii) IP addresses `zmapv6` outputs several times due to potential `SYN ACK` duplicates. While deduplicating IPv4 addresses is comparably easy (`zmap` does it by default), it is not possible to reasonably store information on all IPv6 addresses in memory. Instead, we include (inverse) Bloom filters for the deduplication of IP addresses. Second, to not overload IoT devices with (D)TLS handshakes, we program our IPv4 and IPv6 scanners to wait 15 min between subsequent handshakes to a single deployment. For MQTT, we further limit the connection time (30 min) and outgoing traffic (10 MB) per host.

Load on the Internet: Additionally, to not overload any autonomous system, we limit our scans sending max. 100 k packets per second and randomly order addresses to scan. Additionally, we closely cooperate with our Network Operation Center to handle potential incoming abuse requests.

Contact Information: To give information on the purpose, scope, and (expected) impact of our research, we serve a website that informs about our research and opt-out possibilities on the same IP addresses that we utilize for our Internet scans. We refrain from scanning any "blocklisted" IP addresses again. To date, these blocklisted addresses accumulate to 5.8 M IPv4 addresses and 3.3×10^{30} IPv6 addresses, primarily due to previous scanning activities of our institution. Furthermore, we set up rDNS records for our scanning IP addresses, embed contact information in our client certificate, and include our contact details in protocol messages whenever supported.

B. Validating Responses

After running our scans according to our methodology, we need to validate the results to extract responses that prove the operation of an IoT deployment behind an IP address [19]. To better understand our validation results, we also compare findings from our IPv6 scans to IPv4.

Table II (center) leads through our three-step validation process. For all hosts that respond to our connection attempts (column *Hosts*), we ***first*** filter systems that respond but do not establish a valid TCP connection or answer with faulty UDP packets, e.g., an invalid length field (column *Transport*). While we see similarities across IPv4 and IPv6 scans, fluctuations in the number of answering and filtered IPv6 addresses can be traced to single runs of address generators resulting in many IPs in specific ASes. For example, running 6Forest for OPC UA (port 4843) on v4 results generated more than 1 M IPs in a single AS that all respond but do not establish a valid connection. This result underpins the importance of carefully selecting address generators and their inputs.

Second, we check for deployments that complete a (D)TLS handshake (column *(D)TLS Success*). Notably, on IPv6, similar to IPv4, numerous hosts complete a (D)TLS handshake on the port specified for the non-(D)TLS variant of the respective protocol, already indicating that some IoT deployments run (D)TLS-enabled protocol deployments on the standard port.

Last, we report on deployments that correctly respond to protocol conformant requests (column *Valid*). The large discrepancy between valid IoT-related protocol deployments and successful TCP and (D)TLS handshakes underlines that operators "hide" several non-IoT-related services behind ports intended for IoT protocols. Additionally, some deployments offer TLS on both ports or optional TLS support by providing an insecure and secure endpoint on the respective ports. In the following, we count these deployments only once and as TLS-adopting as they otherwise would distort our analysis. For the IoT-related services under study, we find in total 6658 IPv6-reachable deployments with a strong bias towards protocols usually used for backend services (Backend: 6650, Device: 8). In comparison to IPv4, we find fewer deployments (IPv4: 807 230), indicating both that address generators today do not generate all relevant IP addresses and that probably fewer deployments are reachable via IPv6. Still, the fraction of services that use (D)TLS to secure communication is low: Only 6.2 % of IPv6-reachable services implement (D)TLS (IPv4: 6.3 %) showing that, although implementing a more modern Internet protocol version, only a few deployments and their operators consider security.

C. Information on Valid Deployments

So far, it is unclear in which AS the IoT deployments are located, how many operators run them, and whether they are subject to aliasing. However, this knowledge is required to allow a better understanding of their operation and the influences our measurement methodology might have on the results. Table II (right) lists information on the ASes deployments reside in, common names from received certificates, and the number of IPs that may be subject to aliasing.

Accommodating ASes: Since we found fewer IPv6 deployments, the number of ASes where we found IoT deployments is smaller for IPv6 than for IPv4 (column *ASes-Distinct*). However, the number of ASes per valid deployment is consistently higher. While this view could be distorted by IPv6 measurements not covering the entire address space, it still indicates that found IPv6 deployments are more distributed over the Internet than all IPv4 deployments. Additionally, we found valid IPv6 deployments in ASes where no IPv4 deployment is located (column *ASes-+IPv6*), showing that IPv4-wide studies indeed did not consider all IoT deployments.

Looking at the AS type deployments reside in (column *ASes-Types*[1]), IPv6 deployments, especially MQTT brokers, are significantly more prominent in content-related systems, e.g., cloud networks. Interestingly, seedlists, e.g., from TUM, usually contain significantly more addresses in ISP networks [16]. Thus, this shift indicates that ISP addresses in hitlists might have too short a lifetime, e.g., due to prefix changes when reconnecting, or that operators more likely deploy IPv6-reachable IoT backend services in the cloud.

[1] AS type according to PeeringDB.

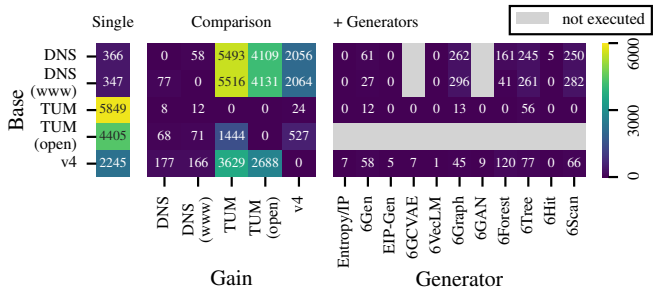

Fig. 1: Comparison of seedlists and number of generated addresses from generators on specific seedlist leading to previously unknown deployments.

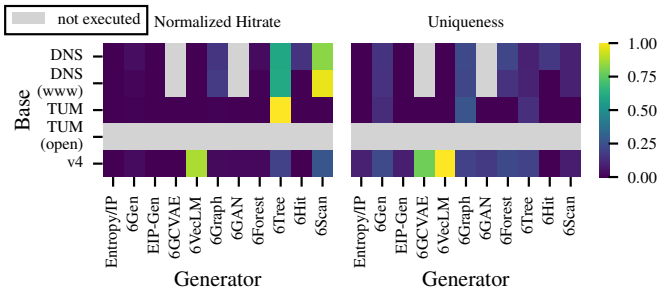

Fig. 2: Normalized hitrate and uniqueness of generators. Addresses often overlap, but 6Scan has a high hitrate.

Operator Information & Aliasing: Similar to the number of ASes per valid deployment, the number of different Common Names per deployment increases as well (column *Cert. CNs-Distinct*), indicating that fewer operators run multiple (D)TLS-enabled deployments. Already this large share of Common Names per deployment suggests that the number of aliased addresses in our dataset can be neglected. The aliasing information offered with the TUM hitlist confirms this presumption (column *Aliasing*). Only 0.7 % of discovered IPv6 deployments are marked as aliased.

Takeaway: *Relying on numerous address sources and generators, we identified 6658 IPv6-reachable IoT deployments with minor subject to aliasing. Notably, and already looking at their security, a similar small fraction as in IPv4 implements TLS to secure their communication (6.2 % vs. 6.3 %).*

IV. Tracing Found Deployments

To understand which address sources helped to find the IPv6-reachable IoT deployments and to guide future studies, we trace found deployments through our generation process.

Seedlists: Figure 1 (left) shows how many IPv6 addresses of valid IoT deployments originate from each address source. Notably, the TUM list leads to the most found deployments and thus constitutes a good starting point for IPv6-wide IoT studies. However, their openly available hitlist (TUM (open)), where all addresses not answering on Web-specific ports or ICMPv6 are filtered, misses IP addresses running IoT-related services, indicating that checking for an IP address's liveness via these protocols is not beneficial and researchers thus should use the unfiltered list for research offside of the Web. Interestingly, also IP addresses out of the DNS lead to IoT deployments. The comparably low gain in finding IoT deployments using DNS in comparison to the TUM list (Figure 1 (center), 8 / 12 (with `www`) deployments), is due to TUM also including forward DNS data, but of other sources [11]. With 24 additionally found deployments, information from IPv4 scans have still a low, but the comparably highest gain in comparison to the TUM list, showing that IoT-related address sources can increase the value of this list.

Generators: To overcome the narrow view of our selected seedlists, we feed them into address generators. Figure 1 (right) shows the number of found IoT services that are found due to generator output but are not part of any seedlist. Most notably, several generators find more active IP addresses using inputs other than the TUM list, indicating that the variety and potentially high age of included addresses do not support current generators. Instead, using input from DNS leads to the most success in generating IPv6 addresses of IoT deployments as most of these IP addresses are most likely currently in use.

To further analyze the effectiveness of the different generators, Figure 2 (left) shows the normalized hitrate, i.e., active IoT-related IP addresses divided by the number of generated IP addresses of each generator, where 1 corresponds to the generator with the highest hitrate (unnormalized: 0.025 ‰). While generating the second-highest number of active addresses, 6Scan on the DNS (www) input has the second-highest hitrate, showing that incorporating information from the Internet during the scan to estimate completely new search directions in addition to the seedlist is beneficial to find IoT deployments. Thus, 6Scan is a promising address generation candidate for future IPv6-wide IoT studies.

Looking at the generators' result sets, we see intersections. Figure 2 (right) shows the *uniqueness* of outputs, defined as the number of addresses found by a generator divided by the sum over the number of all generators generating each found address. Thus, a uniqueness of 1 means that all addresses are only found by this generator. 6GCVAE and 6VecLM with v4 input have a high uniqueness but only helped to find 7 / 1 IoT deployments, making them unique but not effective.

Our results show that generators help to find 768 more IoT deployments than initially included in all seedlists (5890). However, running all generators for IoT-related studies is ineffective due to underperformance, overlapping, and long runtimes. Focussing on a few with comparably high hitrates, i.e., 6Scan and 6Graph on DNS (www), and all available hitlists allows covering 95 % of all our found deployments only with a fraction of time and computing resources.

Takeaway: *While the TUM address list is a good starting point for IPv6-wide IoT studies, generators help to find further active addresses. However, running all generators on all inputs is not required, as their results often overlap.*

V. Security Assessment

Since a *secure* operation is important for IoT deployments as they regularly get in contact with sensitive (user) data, we exemplarily assess the security of IPv6-reachable IoT deployments in comparison to IPv4. In the course of our

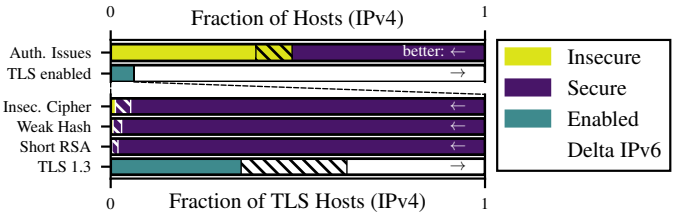

Fig. 3: Security issues of IPv6 deployments in comparison to IPv4. IPv6 installations have similar issues.

assessment, we adhere to previous assessment approaches [19] to maintain further comparability. Specifically, we check for the acceptance of (insecure) (D)TLS cipher suites and try to access openly available data. As we found only a few IPv6-reachable CoAP and OPC UA deployments, we focus our assessment on AMQP and MQTT deployments.

Access Control: Figure 3 (top) shows that fewer of the found IPv6-reachable deployments miss access control in comparison to IPv4 (39 % vs. 48 %). However, the focus of found IPv6 deployments running at cloud providers might bias our result, as these potentially have more security expertise than private users. Still, we found 1930 MQTT and 750 AMQP IPv6 deployments that allow anybody to connect. This negligence might enable attackers to eavesdrop on transmitted IoT data, e.g., from smart homes, or to send malicious commands.

Communication Security: Looking at the 6.2 % of found IPv6-reachable deployments that use TLS to prevent attacks on communications (6.3 % for IPv4), a larger share than in the IPv4 address space fail to adhere to TLS configuration guidelines [66]–[68] (7.8 % vs. 1.8 %). Figure 3 (bottom) shows that a larger share of IPv6-reachable deployments accepts more insecure ciphers and relies on deprecated primitives, e.g., SHA1, or too short RSA keys in their certificates than IPv4 deployments. Thus, attackers can more easily impersonate these deployments to access sensitive data or take over IoT systems.

Considering the newest TLS version, i.e., TLS 1.3, we see a larger fraction of deployments signaling support in comparison to IPv4 (63 % vs. 35 %). This shift suggests that IPv6-reachable deployments are newer and thus can rely on more recent protocol implementations, despite having a penchant for deprecated ciphers and cryptographic primitives.

Takeaway: *Despite the higher adoption of TLS 1.3 at IPv6 IoT systems, suggesting their more recent deployment (63 % vs. 35 % in IPv4), they suffer from similar security issues as their IPv4 counterparts, e.g., no access control (39 % vs. 48 %).*

VI. Discussion & Future Work

The outcome of our work is manifold. While it supports future IoT and other studies with insight into how to scan IPv6, future work could also address its limitations.

Not Covering the Complete Address Space: All IPv6-wide studies, including ours, cannot reliably estimate which portion of installations they covered. Still, by scanning 14 billion addresses, we found 6658 IoT deployments helping to understand an even larger part of the IoT.

For IPv4 scans, depending on the research question, it might be sufficient to randomly scan only 1 % of the address space [69]. However, in the IPv6 Internet, the sparse usage of addresses would lead to skewed results. Instead, it might be beneficial to rely on information gathered from NTP pool servers [70] to increase the low scan-success rate of IPv6 scans (0.000 248 ‰ vs. 0.013 ‰ in IPv4). Additionally, related work can look into the timeliness of IP addresses in seedlists as other works indicate to find more deployments [18].

Address Generator Inputs: Address generators influence the number of found deployments by nudging the scan to specific addresses. Thus, their seeding and configuration might influence the study. While we chose to configure the generators according to the corresponding publications (cf. Section III-A3) and seed them randomly with IPs separated by source, other approaches could produce more active IPv6 addresses. Thus, future work still could look into other seedings, e.g., separated by AS type as done for the Web [16].

Intersecting Installations: While we argue that future IoT studies should include IPv6-reachable deployments to gain a full view, our research leaves open how many IoT deployments are reachable via both IPv6 and IPv4. Especially given that we found 2245 deployments using data from our IPv4 scans as source, the intersection could be high. Still, we found 89 IoT deployments located in ASes that do not accommodate any deployment during our IPv4 scan and thus are, with a high probability, not reachable via IPv4. Future studies could improve the detection of multi-homed deployments by generating fingerprints without considering the IP address.

VII. Conclusion

Internet-wide studies are an indispensable tool to understand how operators manage IoT deployments in the wild, to uncover security flaws, and to derive requirements for mechanisms preventing misconfigurations in the future [19], [20]. However, so far, these studies focused on IPv4-reachable deployments and left out the huge IPv6 address space.

Our results show that not all address generators are beneficial and the seed selection notably influences their result. Two generators and three address lists suffice to detect 95 % of found deployments. Security-wise, we find similar issues in the IPv4 and IPv6 address space: Only 6.2 % of IPv6-reachable deployments implement TLS for communication security (IPv4: 6.3 %) and 39 % fail to implement access control (IPv4: 48 %), enabling attackers to easily access potentially sensitive information.

To conclude, our work shows that the selection of address generators and their seeding is key to extend the findability of IoT devices in the IPv6 address space beyond hitlists. Furthermore, mechanisms targeting to prevent insecure misconfigurations and security assessments must also consider IPv6-reachable deployments.

Acknowledgement

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy — EXC-2023 Internet of Production — 390621612.

References

[1] M. Kassab, J. DeFranco, and P. Laplante, "A systematic literature review on Internet of things in education: Benefits and challenges," *Journal of Computer Assisted Learning*, vol. 36, no. 2, pp. 115–127, 2020.

[2] W. Z. Khan, M. H. Rehman, H. M. Zangoti, M. K. Afzal, N. Armi, and K. Salah, "Industrial internet of things: Recent advances, enabling technologies and open challenges," *Computers & Electrical Engineering*, vol. 81, 2020.

[3] M. Henze, B. Wolters, R. Matzutt, T. Zimmermann, and K. Wehrle, "Distributed Configuration, Authorization and Management in the Cloud-based Internet of Things," in *Proceedings of the 2017 IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom '17)*. IEEE, 2017, pp. 185–192.

[4] F. Maggi, R. Vosseler, and D. Quarta, "The Fragility of Industrial IoT's Data Backbone: Security and Privacy Issues in MQTT and CoAP Protocols," Trend Micro Inc., Tech. Rep., 2018.

[5] M. Piraux, T. Barbette, N. Rybowski, L. Navarre, T. Alfroy, C. Pelsser, F. Michel, and O. Bonaventure, "The Multiple Roles That IPv6 Addresses Can Play in Today's Internet," *ACM SIGCOMM Computer Communication Review*, vol. 52, no. 3, pp. 10–18, 2022.

[6] K. Perset, "Internet Addressing: Measuring Deployment of IPv6," Organisation for Economic Co-operation and Development (OECD), OECD Digital Economy Papers 172, 2010.

[7] M. Dahlmanns, J. Lohmöller, I. B. Fink, J. Pennekamp, K. Wehrle, and M. Henze, "Easing the Conscience with OPC UA: An Internet-Wide Study on Insecure Deployments," in *Proceedings of the ACM Internet Measurement Conference (IMC '20)*. ACM, 2020, pp. 101–110.

[8] Z. Durumeric, F. Li, J. Kasten, J. Amann, J. Beekman, M. Payer, N. Weaver, D. Adrian, V. Paxson, M. Bailey, and J. A. Halderman, "The Matter of Heartbleed," in *Proceedings of the 2014 Conference on Internet Measurement Conference (IMC '14)*. ACM, 2014, pp. 475–488.

[9] M. Nawrocki, P. F. Tehrani, R. Hiesgen, J. Mücke, T. C. Schmidt, and M. Wählisch, "On the Interplay between TLS Certificates and QUIC Performance," in *Proceedings of the 18th International Conference on Emerging Networking EXperiments and Technologies (CoNEXT '22)*. ACM, 2022, pp. 204–213.

[10] D. Adrian, Z. Durumeric, G. Singh, and J. A. Halderman, "Zippier ZMap: Internet-Wide Scanning at 10 Gbps," in *Proceedings of the 8th USENIX Workshop on Offensive Technologies (WOOT '14)*. USENIX Association, 2014.

[11] O. Gasser, Q. Scheitle, P. Foremski, Q. Lone, M. Korczyński, S. D. Strowes, L. Hendriks, and G. Carle, "Clusters in the Expanse: Understanding and Unbiasing IPv6 Hitlists," in *Proceedings of the 2018 Internet Measurement Conference (IMC '18)*. ACM, 2018, pp. 364–378.

[12] J. Zirngibl, L. Steger, P. Sattler, O. Gasser, and G. Carle, "Rusty Clusters? Dusting an IPv6 Research Foundation," in *Proceedings of the 22nd ACM Internet Measurement Conference (IMC '22)*. ACM, 2022, pp. 395–409.

[13] T. Yang, B. Hou, Z. Cai, K. Wu, T. Zhou, and C. Wang, "6Graph: A graph-theoretic approach to address pattern mining for Internet-wide IPv6 scanning," *Computer Networks*, vol. 203, 2022.

[14] G. Song, J. Yang, Z. Wang, L. He, J. Lin, L. Pan, C. Duan, and X. Quan, "DET: Enabling Efficient Probing of IPv6 Active Addresses," *IEEE/ACM Transactions on Networking*, vol. 30, no. 4, pp. 1629–1643, 2022.

[15] B. Hou, Z. Cai, K. Wu, T. Yang, and T. Zhou, "6Scan: A High-Efficiency Dynamic Internet-Wide IPv6 Scanner With Regional Encoding," *IEEE/ACM Transactions on Networking*, 2023.

[16] L. Steger, L. Kuang, J. Zirngibl, G. Carle, and O. Gasser, "Target Acquired? Evaluating Target Generation Algorithms for IPv6," in *Proceedings of the 15th International Workshop on Traffic Monitoring and Analysis (TMA '23)*. IFIP, 2023.

[17] S. J. Saidi, S. Matic, O. Gasser, G. Smaragdakis, and A. Feldmann, "Deep Dive into the IoT Backend Ecosystem," in *Proceedings of the 22nd ACM Internet Measurement Conference (IMC '22)*. ACM, 2022, pp. 488–503.

[18] P. Jose, S. J. Saidi, and O. Gasser, "Analyzing IoT Hosts in the IPv6 Internet," arXiv:2307.09918, 2023.

[19] M. Dahlmanns, J. Lohmöller, J. Pennekamp, J. Bodenhausen, K. Wehrle, and M. Henze, "Missed Opportunities: Measuring the Untapped TLS Support in the Industrial Internet of Things," in *Proceedings of the 17th ACM ASIA Conference on Computer and Communications Security (ASIACCS '22)*. ACM, 2022, pp. 252–266.

[20] S. Srinivasa, J. M. Pedersen, and E. Vasilomanolakis, "Open for Hire: Attack Trends and Misconfiguration Pitfalls of IoT Devices," in *Proceedings of the 21st ACM Internet Measurement Conference (IMC '21)*. ACM, 2021, pp. 195–215.

[21] COMSYS, "IPv6 Scanning Tools," https://github.com/COMSYS/ipv6-scanning, 2024.

[22] T. Chown, "IPv6 Implications for Network Scanning," IETF RFC 5157, 2008.

[23] F. Gont and T. Chown, "Network Reconnaissance in IPv6 Networks," IETF RFC 7707, 2016.

[24] K. Borgolte, S. Hao, T. Fiebig, and G. Vigna, "Enumerating Active IPv6 Hosts for Large-Scale Security Scans via DNSSEC-Signed Reverse Zones," in *Proceedings of the 2018 IEEE Symposium on Security and Privacy (SP '18)*. IEEE, 2018, pp. 770–784.

[25] O. Gasser, Q. Scheitle, S. Gebhard, and G. Carle, "Scanning the IPv6 Internet: Towards a Comprehensive Hitlist," in *Proceedings of the 8th International Workshop on Traffic Monitoring and Analysis (TMA '16)*. IFIP, 2016.

[26] T. Fiebig, K. Borgolte, S. Hao, C. Kruegel, and G. Vigna, "Something from Nothing (There): Collecting Global IPv6 Datasets from DNS," in *Proceedings of the 18th International Conference on Passive and Active Measurement (PAM '17)*, vol. 10176. Springer, 2017, pp. 30–43.

[27] J. Zirngibl, P. Buschmann, P. Sattler, B. Jaeger, J. Aulbach, and G. Carle, "It's over 9000: Analyzing Early QUIC Deployments with the Standardization on the Horizon," in *Proceedings of the 21st ACM Internet Measurement Conference (IMC '21)*. ACM, 2021, pp. 261–275.

[28] X. Li, B. Liu, X. Zheng, H. Duan, Q. Li, and Y. Huang, "Fast IPv6 Network Periphery Discovery and Security Implications," in *Proceedings of the 2021 51st Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN '21)*. IEEE, 2021, pp. 88–100.

[29] G. Song, L. He, Z. Wang, J. Yang, T. Jin, J. Liu, and G. Li, "Towards the Construction of Global IPv6 Hitlist and Efficient Probing of IPv6 Address Space," in *Proceedings of the 2020 IEEE/ACM 28th International Symposium on Quality of Service (IWQoS '20)*. IEEE, 2020.

[30] J. Ullrich, P. Kieseberg, K. Krombholz, and E. Weippl, "On Reconnaissance with IPv6: A Pattern-Based Scanning Approach," in *Proceedings of the 2015 10th International Conference on Availability, Reliability and Security (ARES '15)*. IEEE, 2015, pp. 186–192.

[31] P. Foremski, D. Plonka, and A. Berger, "Entropy/IP: Uncovering Structure in IPv6 Addresses," in *Proceedings of the 2016 ACM Internet Measurement Conference (IMC '16)*. ACM, 2016, pp. 167–181.

[32] A. Murdock, F. Li, P. Bramsen, Z. Durumeric, and V. Paxson, "Target Generation for Internet-Wide IPv6 Scanning," in *Proceedings of the 2017 Internet Measurement Conference (IMC '17)*. ACM, 2017, pp. 242–253.

[33] G. Zheng, X. Xu, and C. Wang, "An Effective Target Address Generation Method for IPv6 Address Scan," in *Proceedings of the 2020 IEEE 6th International Conference on Computer and Communications (ICCC '20)*. IEEE, 2020, pp. 73–77.

[34] T. Cui, G. Gou, and G. Xiong, "6GCVAE: Gated Convolutional Variational Autoencoder for IPv6 Target Generation," in *Proceedings of the 24th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining (PAKDD '20)*, vol. 12084. Springer, 2020, pp. 609–622.

[35] T. Cui, G. Xiong, G. Gou, J. Shi, and W. Xia, "6VecLM: Language Modeling in Vector Space for IPv6 Target Generation," in *Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases: Applied Data Science Track (ECML PKDD '20)*, vol. 12460. Springer, 2021, pp. 192–207.

[36] T. Cui, G. Gou, G. Xiong, C. Liu, P. Fu, and Z. Li, "6GAN: IPv6 Multi-Pattern Target Generation via Generative Adversarial Nets with Reinforcement Learning," in *Proceedings of the 40th IEEE Conference on Computer Communications (INFOCOM '21)*. IEEE, 2021.

[37] T. Yang, Z. Cai, B. Hou, and T. Zhou, "6Forest: An Ensemble Learning-based Approach to Target Generation for Internet-wide IPv6 Scanning," in *Proceedings of the 41st IEEE Conference on Computer Communications (IEEE INFOCOM '22)*. IEEE, 2022, pp. 1679–1688.

[38] Z. Liu, Y. Xiong, X. Liu, W. Xie, and P. Zhu, "6Tree: Efficient dynamic discovery of active addresses in the IPv6 address space," *Computer Networks*, vol. 155, pp. 31–46, 2019.

[39] B. Hou, Z. Cai, K. Wu, J. Su, and Y. Xiong, "6Hit: A Reinforcement Learning-based Approach to Target Generation for Internet-wide IPv6

Scanning," in *Proceedings of the 40th IEEE Conference on Computer Communications (INFOCOM '21)*. IEEE, 2021.
[40] G. Song, J. Yang, L. He, Z. Wang, G. Li, C. Duan, Y. Liu, and Z. Sun, "AddrMiner: A Comprehensive Global Active IPv6 Address Discovery System," in *Proceedings of the 2022 USENIX Annual Technical Conference (ATC '22)*. USENIX Association, 2022, pp. 309–326.
[41] T. Albakour, O. Gasser, R. Beverly, and G. Smaragdakis, "Third Time's Not a Charm: Exploiting SNMPv3 for Router Fingerprinting," in *Proceedings of the 21st ACM Internet Measurement Conference (IMC '21)*. ACM, 2021, pp. 150–164.
[42] M. Luckie, R. Beverly, W. Brinkmeyer, and k. claffy, "Speedtrap: Internet-Scale IPv6 Alias Resolution," in *Proceedings of the 2013 Internet Measurement Conference (IMC '13)*. ACM, 2013, pp. 119–126.
[43] R. Beverly, W. Brinkmeyer, M. Luckie, and J. P. Rohrer, "IPv6 Alias Resolution via Induced Fragmentation," in *Proceedings of the 14th International Conference on Passive and Active Measurement (PAM '13)*, vol. 7799. Springer, 2013, pp. 155–165.
[44] K. Vermeulen, B. Ljuma, V. Addanki, M. Gouel, O. Fourmaux, T. Friedman, and R. Rejaie, "Alias Resolution Based on ICMP Rate Limiting," in *Proceedings of the 21st International Conference Passive and Active Measurement (PAM '20)*, vol. 12048. Springer, 2020, pp. 231–248.
[45] A. Marder, "APPLE: Alias Pruning by Path Length Estimation," in *Proceedings of the 21st International Conference Passive and Active Measurement (PAM '20)*, vol. 12048. Springer, 2020, pp. 249–263.
[46] R. Padmanabhan, Z. Li, D. Levin, and N. Spring, "UAv6: Alias Resolution in IPv6 Using Unused Addresses," in *Proceedings of the 16th International Conference on Passive and Active Measurement (PAM '15)*, vol. 8995. Springer, 2015, pp. 136–148.
[47] T. Albakour, O. Gasser, and G. Smaragdakis, "Pushing Alias Resolution to the Limit," in *Proceedings of the 2023 ACM Internet Measurement Conference (IMC '23)*, Oct. 2023.
[48] Z. Durumeric, D. Adrian, A. Mirian, M. Bailey, and J. A. Halderman, "A Search Engine Backed by Internet-Wide Scanning," in *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security (CCS '15)*. ACM, 2015, pp. 542–553.
[49] Censys, "Censys," https://www.censys.io, 2023.
[50] É. P. Leverett, "Quantitatively Assessing and Visualising Industrial System Attack Surfaces," Master's thesis, University of Cambridge, 2011.
[51] A. Hansson, M. Khodari, and A. Gurtov, "Analyzing Internet-connected industrial equipment," in *Proceedings of the 2018 International Conference on Signals and Systems (ICSigSys '18)*. IEEE, 2018, pp. 29–35.
[52] H. Al-Alami, A. Hadi, and H. Al-Bahadili, "Vulnerability scanning of IoT devices in Jordan using Shodan," in *Proceedings of the 2017 2nd International Conference on the Applications of Information Technology in Developing Renewable Energy Processes & Systems (IT-DREPS '17)*. IEEE, 2017.
[53] B. Zhao, S. Ji, W.-H. Lee, C. Lin, H. Weng, J. Wu, P. Zhou, L. Fang, and R. Beyah, "A Large-Scale Empirical Study on the Vulnerability of Deployed IoT Devices," *IEEE Transactions on Dependable and Secure Computing*, vol. 19, no. 3, pp. 1826–1840, 2022.
[54] J. M. Ceron, J. J. Chromik, J. Santanna, and A. Pras, "Online Discoverability and Vulnerabilities of ICS/SCADA Devices in the Netherlands," arXiv:2011.02019, 2020.
[55] T. Kiravuo, S. Tiilikainen, M. Särelä, and J. Manner, "Peeking Under the Skirts of a Nation: Finding ICS Vulnerabilities in the Critical Digital Infrastructure," in *Proceedings of the 14th European Conference on Cyber Warfare and Security (ECCWS '15)*. ACPI, 2015, pp. 137–144.
[56] B. Genge and C. Enăchescu, "ShoVAT: Shodan-Based Vulnerability Assessment Tool for Internet-Facing Services," *Security and Communication Networks*, vol. 9, no. 15, pp. 2696–2714, 2016.
[57] G. Barbieri, M. Conti, N. O. Tippenhauer, and F. Turrin, "Assessing the Use of Insecure ICS Protocols via IXP Network Traffic Analysis," in *Proceedings of the 2021 International Conference on Computer Communications and Networks (ICCCN '21)*. IEEE, 2021.
[58] K. Yang, Q. Li, and L. Sun, "Towards automatic fingerprinting of IoT devices in the cyberspace," *Computer Networks*, vol. 148, pp. 318–327, 2019.
[59] Z. Durumeric, E. Wustrow, and J. A. Halderman, "ZMap: Fast Internet-wide Scanning and Its Security Applications," in *Proceedings of the 22nd USENIX Security Symposium (SEC '14)*. USENIX Association, 2013, pp. 605–620.
[60] Shodan, "Shodan," https://www.shodan.io, 2013.
[61] Y. Wu, C. Li, J. Yang, Z. Wang, W. Hu, A. Xia, Y. Jiang, Y. Wang, and L. Wu, "WebIoT: Classifying Internet of Things Devices at Internet Scale through Web Characteristics," in *Proceedings of the 2022 IEEE Symposium on Computers and Communications (ISCC '22)*. IEEE, 2022, pp. 1–7.
[62] Internet Corporation for Assigned Names and Numbers, "Centralized Zone Data Service," https://czds.icann.org/home, 2020.
[63] TU Munich – Chair of Network Architectures and Services, "ZMapv6: Internet Scanner with IPv6 capabilities," https://github.com/tumi8/zmap, 2021.
[64] COMSYS, "ZGrab 2.0," https://github.com/COMSYS/zgrab2, 2021.
[65] D. Dittrich and E. Kenneally, "The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research," U.S. Department of Homeland Security, Tech. Rep., 2012.
[66] K. A. McKay and D. A. Cooper, "Guidelines for the Selection, Configuration, and Use of Transport Layer Security (TLS) Implementations," NIST SP 800-52 Rev. 2, 2019.
[67] Y. Sheffer, R. Holz, and P. Saint-Andre, "Recommendations for Secure Use of Transport Layer Security (TLS) and Datagram Transport Layer Security (DTLS)," IETF RFC 7525, 2015.
[68] Federal Office for Information Security, "Cryptographic Mechanisms: Recommendations and Key Lengths: Use of Transport Layer Security (TLS)," BSI TR-02102-2, 2021.
[69] J. Rüth, I. Kunze, and O. Hohlfeld, "TCP's Initial Window—Deployment in the Wild and Its Impact on Performance," *IEEE Transactions on Network and Service Management*, vol. 16, no. 2, pp. 389–402, 2019.
[70] E. Rye and D. Levin, "IPv6 Hitlists at Scale: Be Careful What You Wish For," in *Proceedings of the 2023 ACM SIGCOMM Conference (SIGCOMM '23)*. ACM, 2023, pp. 904–916.