6Seeks: A Global IPv6 Network Periphery Scanning System

Tao Yang, Bingnan Hou, Yifan Yang, Zhenzhong Yang, and Zhiping Cai

***Abstract*—Discovering the IPv6 network periphery, i.e., the last-hop router connecting endhosts in the IPv6 Internet, is crucial for network measurement and Internet reconnaissance. However, existing solutions commonly suffer from inefficiency when applied on a global scale due to the vast IPv6 address space. To tackle this challenge, we developed *6Seeks*, an innovative IPv6 scanning system designed for efficient IPv6 periphery discovery across the global IPv6 Internet without requiring seed IPv6 addresses. Specifically, we proposed to employ a heuristic method for collecting active /48 networks from the global BGP prefixes and then adopt a reinforcement learning-based dynamic probing strategy to optimize resource allocation across these networks and significantly improve efficiency. Real-world tests demonstrate that *6Seeks* outperforms all existing methods in global-scale IPv6 periphery measurement experiments. In just a few hours, *6Seeks* can identify over 128 million IPv6 periphery devices, while using only 37% of the probing resources required by the current state-of-the-art solution. Compared to existing public datasets, the IPv6 addresses identified by *6Seeks* are more numerous and display unique characteristics, significantly enriching our IPv6 corpus.**

***Index Terms*—IPv6, internet-wide scanning, network measurement.**

I. Introduction

THE vast IPv6 address space revolutionizes Internet addressing principles, enabling end-users to obtain one or more addressable IPv6 prefixes instead of relying on a single private IPv4 address assigned via Network Address Translation (NAT) [1], [2]. This advancement facilitates direct end-to-end communication between individual IPv6 devices and Internet. The *IPv6 periphery*, typically the last-hop router connecting end-hosts, is responsible for packet forwarding [3] and traffic filtering [4]. Additionally, these devices serve as gateways, handling access control, service hosting, system provisioning, and other critical functions [5]. Consequently, the IPv6 periphery, as an essential part of the Internet architecture, plays a critical role in maintaining the stability and security of the IPv6 Internet.

Received 24 June 2024; revised 20 December 2024; accepted 25 May 2025; approved by IEEE TRANSACTIONS ON NETWORKING Editor D. Malone. Date of publication 17 June 2025; date of current version 17 October 2025. This work was supported in part by the National Natural Science Foundation of China under Grant 62472434 and in part by the Science and Technology Innovation Program of Hunan Province under Grant 2022RC3061. *(Corresponding author: Zhiping Cai.)*

The authors are with the College of Computer Science and Technology, National University of Defense Technology, Changsha 410073, China (e-mail: yangtao97@nudt.edu.cn; houbingnan19@nudt.edu.cn; yangyifanyyf@nudt.edu.cn; zzy.nudt@nudt.edu.cn; zpcai@nudt.edu.cn).

Digital Object Identifier 10.1109/TON.2025.3575443

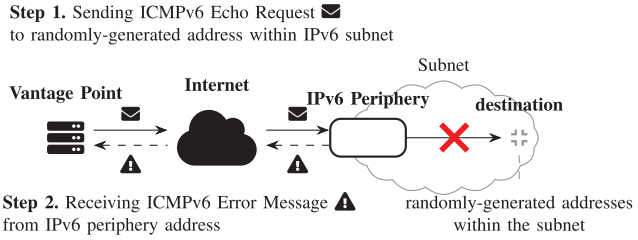


Fig. 1. The paradigm for IPv6 network periphery discovery. The prober at the vantage point sends the IPv6 probes (e.g., ICMPv6 Echo Requests) to randomly-generated targets within the endpoint subnet. Since these probes hardly encounter active hosts, the IPv6 periphery devices instead return ICMPv6 error messages, which disclose their own addresses [12].

Efficient IPv6 periphery discovery can not only provide a valuable IPv6 address corpus but also support a range of security applications, such as network asset censuses [6], privacy leakage detection [7], IP geolocation [8], network reconnaissance [9], [10], and remote network measurement via side channels [11]. As such, there is a compelling incentive to advance research efforts in IPv6 periphery discovery.

Existing solutions for IPv6 network periphery discovery, such as *Xmap* [13] and *Edgy* [14], commonly adopt the paradigm shown in Figure 1, i.e., probing the inactive addresses across specific subnets in order to provoke *indirect responses* from the IPv6 periphery. In other words, we need not know the exact addresses of the IPv6 periphery device responsible for the host subnet in advance. However, directly scaling this paradigm to a global scale is challenging due to the immense IPv6 address space. As outlined in established best practices [15], an IPv6 end-site must be assigned at least one /64 subnet. This allocation implies that the IPv6 Internet theoretically comprises up to 2^{64} possible /64 subnets. Even with only one probe per /64 subnet, a comprehensive scan would still require a prohibitive number of 2^{64} packets. Exhaustively probing these subnets using existing methods across such an expansive range is highly ineffective, analogous to searching for a "needle in a haystack." Thus, discovering the IPv6 network periphery across the global Internet remains an open challenge.

Over the past decade, reinforcement learning (RL) algorithms have been widely used in active IPv6 address detection [16], [17], [18], [19] and IPv6 topology discovery [20], [21]. Related studies [22], [23] have shown that adaptively refining probing strategies based on prior scan results can significantly improve the performance of active IPv6 measurement. This

prompts the inquiry: *Can reinforcement learning algorithms enhance the efficiency of the IPv6 network periphery discovery?* Our answer is positive.

To this end, we developed a reinforcement learning-based IPv6 network periphery scanning system, *6Seeks*, which revolves around the fundamental insight: to maximize efficiency, probing efforts should focus on IPv6 address subspaces with higher returns. Specifically, it divides the global IPv6 network periphery discovery campaign into two phases: *candidate region collection* and *dynamic region probing*.

First, *6Seeks* thoroughly examines the constituent /48 networks within global BGP prefixes, and applies heuristics to identify those that warrant further probing, denoted to the *candidate regions* (see § III). *6Seeks* operates solely using BGP prefixes, which are readily accessible through Looking Glass services [24], [25], [26] or the RouteViews project [27]. This flexibility enable the easy replication of our experiments without requiring IPv6 seed addresses, thereby avoiding the measurement inconsistencies caused by seed biases.

Second, *6Seeks* dynamically probes these candidate regions rather than exhaustively testing all potential targets. It strategically adjusts the allocated probing budget – the number of allowed probing attempts – for each region based on feedback from previous probing outcomes. Within each region, *6Seeks* employs a quasirandom probing sequence to ensure an even distribution of probes. Consequently, regions yielding higher returns receive additional probing budgets while those with the lower returns are allocated fewer throughout. This approach embodies the principles of reinforcement learning. More considerations and details about *6Seeks*'s implementation will be disclosed in § IV.

At its core, *6Seeks* can conserve probes that would otherwise be wasted, significantly reducing the resources required for global IPv6 network periphery discovery. To substantiate this claim, we conducted the global-scale measurement experiments involving *6Seeks* and existing methods on real-world networks. Our findings reveal that *6Seeks* can identify over 128 million unique last-hop router addresses (IPv6 periphery addresses) using only 1.40 billion probes, while the current state-of-the-art approach (*Edgy* [14]) required 3.78 billion probes to achieve similar results, realizing a 64% reduction in resource consumption.

We present a comprehensive analysis comparing the address sets obtained from our IPv6 network periphery measurements with publicly available IPv6 address repositories [18], [19], [28]. The results demonstrate that the *6Seeks*'s dataset significantly enriches our IPv6 corpus. Specifically:

- Our dataset includes 116.4 million /64 prefixes, which are completely disjoint from existing repositories. (§ V-B1)
- Our dataset exhibit distinctive distributions across autonomous systems. (§ V-B2)
- Our dataset reveals unique addressing patterns, particularly a substantial presence of 61.4 million EUI-64 IPv6 addresses which directly expose the devices' MAC addresses to open Internet [29]. (§ V-B3)
- Our dataset has a negligible ratio of IPv6 aliased addresses, ensuring high data reliability. (§ V-B4)

Additionally, our global IPv6 network periphery measurements revealed numerous new discoveries. In summary, the paper's contributions are as follows:

- We proposed a reinforcement learning methodology for optimizing probe resource allocation in global IPv6 network periphery discovery. This is the first application of reinforcement learning in this field.
- We designed and implemented *6Seeks*, an innovative IPv6 scanning system based on the above methodology, whose source code is available at https://6Seeks.github.io.
- We conducted the first global IPv6 network periphery measurement using *6Seeks*, discovering 128.3 million last-hop IPv6 router addresses from a single vantage point with just 37% of the probing budget required by the current state-of-the-art approach.
- We comprehensively compared *6Seeks*'s discoveries with existing IPv6 address repositories, finding that it not only contains the largest number of IPv6 addresses but also exhibits unique properties in many aspects, such as address space coverage, AS-level distribution, addressing patterns, and aliased ratios, highlighting its significant complementarity to existing IPv6 corpora.
- We revealed that approximately 50% of IPv6 periphery devices worldwide still use the legacy EUI-64 Stateless Address Autoconfiguration (SLAAC) addressing [29], which is vulnerable to device tracking attacks.
- We found that, for last-hop IPv6 routers, the 64 least-significant bits of their addresses – the Interface Identifier (IID) – may be associated with their address lifespan. Specifically, the lower the normalized Shannon entropy of the IIDs, the more likely the addresses are to be static and have a long lifespan (e.g., greater than 30 days).

II. Background

A. Related Works

Recently, numerous remarkable studies have been conducted in related domains. To facilitate illustration, we summarize the key aspects of global-scale IPv6 network periphery discovery.

Motivations. Both active IPv6 address detection [30], [31], [32], [33], [34], [35] and network asset censuses [9], [36], [37] necessitate a large and comprehensive IPv6 address set for initial seeds of target generation algorithms (TGAs) and destinations of probing, respectively. References [7], [38], and [39] employ IPv6 periphery addresses using legacy EUI-64 addressing scheme to analyze privacy leakage and cross-network tracking risks of hosts. Efficient discovery of IPv6 periphery addresses provides a high-quality corpus essential for these works; in addition, IPv6 periphery devices can serve not only as landmarkers for unprivileged street-level IPv6 geolocation [8] but also as traffic reflectors for accomplishing remote network measurements via side-channel information [11]. Naturally, conducting the global IPv6 network periphery discovery would enhance the effectiveness of these methodologies.

Challenges. *Xmap* [13] exhaustively scanned 15 IPv6 ISP network ranges to trigger ICMPv6 error messages at last-hop IPv6 routers (the IPv6 periphery devices) collecting 52

million addresses. *Edgy* [14] extracted seed /48 networks from CAIDA topology dataset [40] and iteratively examined the subneting prefixes, intentionally triggering last-hop's indirect responses to collect 64 million IPv6 periphery addresses. However, applying these methods on a global scale would be challenging due to the prohibitive search space: *Xmap* struggles with the immense volume of global BGP prefixes, and *Edgy* lacks sufficient seed /48 networks as the input. We will show that straightforwardly using these approaches for global-scale measurement campaign will suffer significant inefficiency.

Reinforcement learning in IPv6 scanning. *6Hit* [16] pioneeringly applied reinforcement learning (RL) to dynamically scan active IPv6 addresses, mitigating biases from low-quality initial seeds and significantly outperforming prior methods. *AddrMiner* [19] employed Thompson Sampling to detect active addresses in seedless IPv6 regions. *6Sense* [36] used reinforcement learning coupled with an online scanner to iteratively reduce the search space of possible IPv6 addresses. Besides, *Treestrace* [20] integrated reinforcement learning with Huffman coding for dynamic IPv6 topology discovery, while *Sweeper* [21] further optimized budget allocation efficiency to significantly reduce overhead in probing direction adjustments. As a result, both approaches discovered significantly more IPv6 router interfaces in real-world networks. In summary, integrating IPv6 active measurement with reinforcement learning algorithm proved to offer substantial efficiency gains, and we thus aim to apply it for promoting the IPv6 network periphery discovery.

B. Terminology

We borrow the definition of *IPv6 network periphery* from [13] and [14], and outline some related terminologies, referring to [41].

Customer prefix allocation. In IPv6, a customer prefix defines the IPv6 address range that an end-site can use within its LAN, a.k.a., customer network. IPv6 network periphery functions as the last-hop routing gateway device, providing a connection between the Internet and hosts (including itself) of the LAN. RFC6177 [15] recommends that providers allocate prefixes ranging from /48 to /64 for end-sites, but leaves the decision up to individual service providers. This flexibility means that a /48 network can be used by a provider to allocate /64 subnets to 65,536 customers (2^{64-48}), or /56 subnets to 256 customers (2^{64-56}), or a single /48 subnet to only one customer, etc.

Regions and exploration values. Implementing the reinforcement learning-based dynamic probing logic in this paper requires partitioning the IPv6 address space to support budget adjustments. As illustrated in Figure 2, the /48 prefix represents the smallest globally routable BGP prefix size. The prefixes larger than /48 typically belong to provider networks, while /48 or smaller prefixes are generally used for connectivity of customer networks. Accordingly, we define a /48 prefix as a *region*, and its *exploration value* is given by the total count of new last-hop response addresses potentially received when probing within this /48. For instance, if an entire region is allocated to an end site, its exploration value might be initially 1. On the opposite end of the spectrum, a /48 network could have an initial exploration value of 2^{16} if each /64 serves a distinct customer subnet. If we have already discovered all the last-hop IPv6 router addresses within a region, its exploration value will have been exhausted. Probing the regions with no exploration value will cause the waste of resources, and we are motivated to prevent this.

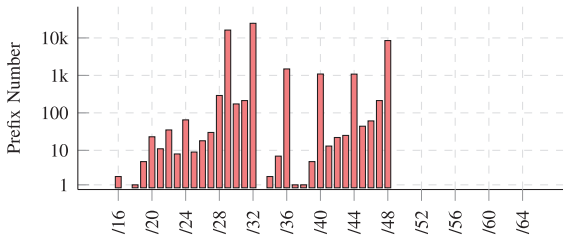


Fig. 2. The distribution of global IPv6 BGP prefixes in the RouteViews project after deduplication and aggregation. Note that no BGP prefix has a longer length than the /48.

TABLE I
IPv6 INTERFACE IDENTIFIER (IID) CATEGARIZATION

Categary	IID Example	Comment
EUI-64	`0250:56ff:fe89:49be`	hardware MAC `00:50:56:89:49:be`
Randomized	`10de:51e8:eb66:7583`	pseudorandom
Embed-IPv4	`0012:0122:0126:0072`	IPv4 address `12.122.126.72` is embedded
Embed-Port	`0000:0000:0000:0080`	TCP/UDP service port `80` is embedded
ISATAP	`0000:5efe:c0a8:0001`	`0000:5EFE` + hexadecimal `192.168.0.1`
Low-byte	`0000:0000:0000:f1b7`	all zeros except the lower bytes
Byte-pattern	`0021:2222:0001:0001`	more than two bytes of zeros

Active and inactive networks/regions. When a network prefix appears in routing tables, packets sent to its addresses can be routed toward designated networks. On the the receiver's network, the last-hop IPv6 router (IPv6 periphery device) would forward packets to their destinations by conducting the Neighbor Discovery [42] to resolve IP addresses into link-layer addresses, thereby accomplishing packet delivery. In this study, we label such networks as *active networks*. Conversely, if the last-hop router is either absent or discards traffic, the network is considered *inactive*. Similarly, an active /48 network is termed an *active region*, while an inactive one is an *inactive region*. The expected responsive IPv6 periphery addresses exist solely in active networks.

Categories and normalized Shannon entropy in IIDs. A 128-bit IPv6 unicast address comprises a 64-bit *network prefix* and a 64-bit *interface identifier* (IID). RFC7707 [10] categorizes these IIDs based on real-world IPv6 address assignment scenarios, as summarized in Table I. EUI-64 IIDs, for instance, are constructed by modifying the MAC address: the 7-th least significant bit is inverted, and `0xFFFE` is inserted between the third and fourth bytes to form a 64-bit identifier [29]. Identifying these EUI-64 IIDs is straightforward – **the 4-th and 5-th bytes are always `0xFF` and `0xFE`, respectively**. Thus, the IPv6 address using EUI-64 IID can easily expose its device hardware MAC address, thereby warranting separate examination.

However, some specific IIDs are difficult to identify without prior knowledge. For example, without additional probing efforts, distinguishing between Embedded-Port and Low-byte

IIDs, as well as verifying Embedded-IPv4 IIDs and their corresponding IPv4 addresses, is nearly impossible. Building upon established works [28], [39], **we use normalized Shannon entropy as a metric for characterizing IPv6 Interface Identifiers (IIDs) in this paper**. Specifically, we model the IID as a sequence of 32 nybbles, where each nybble represents a hexadecimal character. Let X_j denote a discrete random variable over the sample space $\Omega = \{0, 1, \cdots, f\}$, corresponding to the j-th nybble. The empirical probability mass function of X_j is given by $P(X_j)$. The normalized Shannon entropy is then computed as follows:

$$H = -\sum_{\omega \in \Omega} P\left(X_j = w\right) \log_{16}\left(P\left(X_j = w\right)\right)$$

Direct and indirect responses. Given the extremely low probability of encountering a single active address within a /64 prefix ($\frac{1}{2^{64}}$), we consider IPv6 addresses with randomly-generated IIDs to be highly likely *inactive*. Probing inactive addresses within an active network typically elicits ICMPv6 error messages from the last-hop IPv6 router (IPv6 periphery device), indicating that the destination is unreachable. Since the destination address of the probe differs from the source address of the response, we term these responses *indirect*. However, probes receive direct responses, such as ICMPv6 echo replies, from the targets with randomly-generated IIDs, e.g., within IPv6 aliased prefixes. As these direct responses cannot be attributed to the last-hop routers, they are excluded from our IPv6 network periphery discovery results. Additionally, a customer subnet (e.g., a /56) could include a group of /64 prefixes, and probing any of them may receive the homogeneous *indirect* responses from the same last-hop IPv6 router. Hence, we should avoid sending probes to the same customer subnet to prevent wasting resources.

III. Phase 1: Candidate Region Collection

6Seeks uses global BGP prefixes[1] as the starting point for IPv6 network periphery discovery. The RouteViews global routing system [27] provides approximately 58,000 BGP prefixes after deduplication and aggregation, with their length distribution shown in Figure 2. To support *6Seeks*'s probing logic, we partition the IPv6 address space within these BGP prefixes into a series of subspaces, ultimately producing a total of 10.4 billion regions (a.k.a., /48 networks). However, exhaustively probing all 10.4 million regions is neither feasible nor necessary. Instead, the first phase of *6Seeks* focuses on eliminating those inactive regions early, thereby isolating the *potentially* active ones as candidate targets for further probing.

As discussed in § II-B, sending probes to active /48 networks will receive the *indirect* responses from the IPv6 periphery devices of respective subnets. By contrast, probes directed to inactive /48 networks are usually dropped by intermediate-hop routers because their destinations are unreachable. As such, the indirect responses from active /48 networks are typically heterogeneous, while the indirect responses from inactive /48 networks can usually be aggregated based on their source addresses. Based on these observations, we adopt a heuristic method to identify candidate regions: We randomly select a /64 subnet per /48 region and append a random IID to create a target address set equivalent to 10.5 billion regions (**Step 1**). We then send probes to these targets with a *Hop Limit* set to the maximum of 255 and record the indirect responses (**Step 2**). Finally, we retain these responses if their source addresses are unique and label the corresponding destination /48 networks as *potentially* active (**Step 3**). Results show that the entire heuristic-based probing campaign, conducted at an uplink speed of 50 kpps, spanned less than 60 hours and identified 4,473,239 /48 networks as *potentially*-active ones worldwide. These findings will serve as the candidate regions in the next phase of probing.

Admittedly, this heuristic-based approach may occasionally result in the omission of certain regions, i.e., misclassifying active regions as inactive ones if our probes are directed toward unused portions of the address space within an active /48 network. Nonetheless, this limitation will not compromise the overall effectiveness of *6Seeks* itself, for two primary reasons: 1) such omissions can be effectively mitigated through the simple repetition of the candidate region collection process, and 2) we will show that, even in the presence of such omissions, *6Seeks* retains the capability to efficiently identify IPv6 periphery addresses on a global scale.

IV. Phase 2: Dynamic Region Probing

We have already collected the potentially-active regions from global BGP prefixes as candidate regions, as described in § III. This section details the second phase of *6Seeks*, where we perform dynamic probing on these candidate IPv6 regions, as shown in Figure 3.

To achieve this, we are committed to developing a novel asynchronous IPv6 scanner. It not only achieves high parallelism in probing by decoupling sending and receiving tasks but also incorporates the ability to optimize the allocation of probing resource based on reinforcement learning principles. To clarify its implementation, we first introduce the *probe construction and state control* mechanisms, followed by a detailed examination of its key technical components: *reinforcement learning-based probing strategy* and *quasirandom probing sequence*. This scanner is written in Go, exclusively using the standard library, and is compatible with various UNIX-like platforms. We make its source code publicly available, along with the data collected in our study.

A. Probe Construction & State Control

To ensure broad compatibility across diverse network environments, the probe used in the *6Seeks* scanner essentially resembles a standard ICMPv6 Echo Request. The ICMPv6 protocol is designed for diagnostic purposes and is less intrusive [14]. However, *6Seeks* customizes the probe packets to achieve its measurement objectives. Specifically, we configure the Hop Limit to the maximum value of 255 to ensure target network reachability and populate the least-significant 64 bits of the destination address with randomly-generated Interface Identifiers (IIDs) to avoid encountering active hosts.

[1]Data used in this paper is available at https://archive.routeviews.org/routeviews6/bgpdata

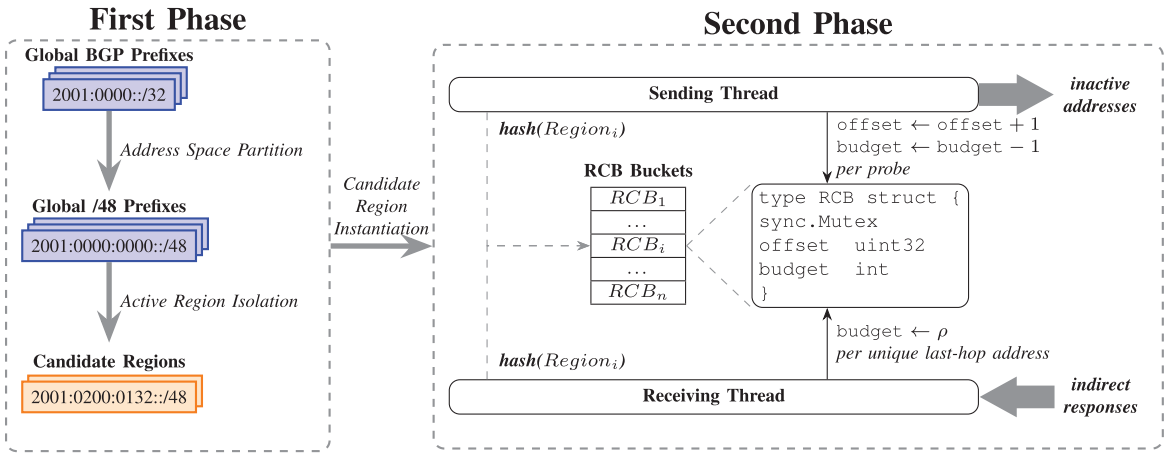


Fig. 3. The *6Seeks*'s workflow for the global IPv6 network periphery measurement consists of two distinct phases.

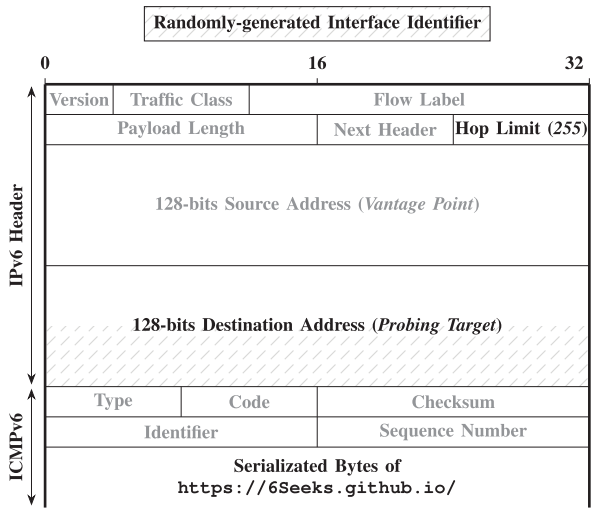


Fig. 4. An example of the probe packet in *6Seeks* scanning system.

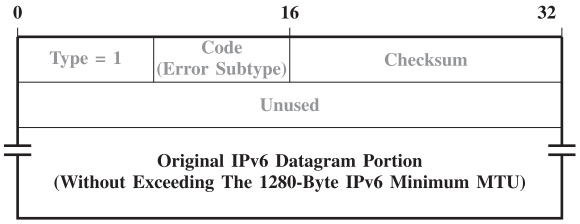


Fig. 5. An example of the ICMPv6 Destination Unreachable message. In an indirect response, its original IPv6 datagram portion typically contains the information of our probe packets.

Additionally, the probe packet's payload exclusively includes a website, as shown in Figure 4, which clarifies the benign intent of our measurement campaigns.

Additionally, *6Seeks* must manage regional intermediate states to support its reinforcement learning-based probing logic. However, the specifics of this state control differ significantly from existing approaches. For each candidate region, *6Seeks* instantiates a "Region Control Block" (*RCB*) structure, shown in Figure 3, which includes a mutual exclusion lock and two integer variables. The lock ensures that a single thread accesses each RCB instance at a time, preventing race conditions that could result in inconsistent states or data corruption. The variables `offset` and `budget` record the quantities of the probes sent and the number of allowed probing attempts remaining within the corresponding region, respectively.

In the asynchronous *6Seeks* scanning system, the decoupling of sending and receiving tasks necessitates that both threads are able to quickly locate the relevant "Region Control Block" (RCB) for any given region. To achieve this, we map each regional /48 prefix to its corresponding RCB using a hash table. While these mappings are straightforward when sending probes, they become significantly more complex upon the receipt of indirect responses. The core challenge lies in accurately identifying the regions (/48 prefixes) associated with the received indirect responses. RFC4443 [12] mandates that:

Every ICMPv6 error message (type <128) MUST include as much of the IPv6 offending (invoking) packet (the packet that caused the error) as possible without making the error message packet exceed the minimum IPv6 MTU.

This stipulation allows us to extract not only the last-hop router address from the IPv6 header of an indirect response but also the probe's destination address from the accompanying ICMPv6 error messages, as shown in Figure 5. As such, we truncate the 48 first-significant bits of the probe destination address to obtain the corresponding regional /48 prefix, instead of explicitly encoding the regional /48 prefix into the payload of probe packets.

The entire structure of RCB buckets, including the associated overhead such as the hash map, consumes 300 MB of memory across all candidate regions, a figure well within the capacity of modern computers. There is a potential to furthe reduce this memory footprint, primarily by replacing the general per-RCB mutexes with more efficient atomic operations. We are planning to implement this optimization in the near future. To conclude, the *6Seeks* system functions effectively without requiring additional investment in costly hardware resources.

Algorithm 1 Dynamic Region Probing Algorithm

Require: P, set of candidate /48 networks.

repeat
  for all $p \in P$ **do**
    if $\texttt{budget}(p) > 0$ **then**
      $\texttt{offset}(p) \leftarrow \texttt{offset}(p) + 1$
      $\texttt{budget}(p) \leftarrow \texttt{budget}(p) - 1$ # replenished per unique last-hop address
      probe(p) # quasirandom probing sequence in § IV-C
    else
      $P \leftarrow P \setminus \{p\}$ # eliminate valueless /48 networks
    end if
  end for
until $P \neq \emptyset$

B. Reinforcement Learning-Based Probing Strategy

Discovering the IPv6 network periphery by exhaustively testing all /64 prefixes in candidate regions will pose prohibitive probe consumption. To address this, a direct approach is to focus probing efforts on those candidate regions with high exploration values. However, the practicality of this approach is limited due to the lack of prior knowledge about customer prefix allocations of those candidate regions.

Existing reinforcement learning-based approaches for active IPv6 address detection usually involves dividing the entire measurement campaign into iterative rounds, and leverages the insights from earlier probing rounds, such as historical hit rates,[2] to optimize resource allocation. As a result, the IPv6 address *subspaces* exhibiting high hit rates historically are prioritized for the following budget allocation, while those performing poorly receive less probes in subsequent measurement stages. However, directly applying this method to the IPv6 network periphery discovery may not be effective.

Specifically, for an active /48 network (candidate region), **its historical performance might not accurately reflect the exploration value of further probing it**. For example, while probing different /64 prefixes within a single customer subnet usually provokes indirect responses from the same last-hop router, only the first probing attempt provides meaningful information; subsequent attempts are often redundant. In practice, it becomes increasingly challenging to uncover new IPv6 periphery addresses as probing progresses. As a result, relying on historical performance for budget allocation like existing methods can easily dispatch too many probes to regions that have exhausted their exploration value, thereby causing inefficiencies. In other words, the crux to addressing this issue is promptly stopping probing those regions with exploration values exhausted.

To achieve this, *6Seeks* introduces a variable called `budget`, which functions as a pool with a finite capacity ρ. This pool is used to manage the number of remaining probing attempts for a specific candidate region. As shown in the Algorithm 1, each probe sent from within a region decreases the `budget` by one, and the `budget` is replenished only when a new IPv6 last-hop router address is discovered. This approach ensures that regions with high exploration values receive sufficient budget due to frequent discovery of new IPv6 periphery addresses. In contrast, regions whose exploration values have been exhausted cannot refill their `budget` pools and quickly deplete their original budget (less than ρ), thus being excluded from subsequent measurement campaigns to avoid the waste of probing resources.

In the probing strategy described, the capacity of the `budget` pool, denoted as ρ, is crucial in regulating the probe usage. Essentially, ρ represents the maximum number of probes allowed before a region is expected to yield a new IPv6 last-hop router address. Setting ρ too high can lead to excessive probes being consumed on candidate regions with limited exploration potential. Conversely, setting it too low may cause premature convergence on a suboptimal subset, risking the neglect of the regions with high exploration values. To strike the right balance, we empirically set ρ to 64, as discussed in § V-A2. This value ensures that probes are focused on genuinely high-value regions while simultaneously eliminating unpromising candidate regions early, thereby minimizing unnecessary resource waste.

C. Quasirandom Probing Sequence

6Seeks leverages prior probing feedback to decide whether to further explore specific candidate regions. To ensure accurate assessment of a region's exploration value, it is essential to spread the target /64 prefixes evenly across the regional IPv6 address space. Otherwise, probing feedbacks can be severely distorted, leading to inefficient allocation of probing resources. For example, as illustrated in Figure 6, an inappropriate probing sequence may concentrate traffic within a single customer subnet, thereby neglecting others. This practice will not only waste the probing resources but also create the false impression that the region has been fully explored, ultimately causing the loss of IPv6 periphery addresses.

To address this challenge, we propose a simple yet effective approach, *quasirandom probing sequence*, for quickly and evenly distributing target /64 prefixes within a candidate region. Specifically, we introduce a variable `offset` for each candidate region, which increments by one per probe. Then, we calculate its bit-reversed value to fill the bits between the 48th and 64th positions, enabling efficient target /64 prefix generation for probing. This method effectively prevents the over-concentration of /64 prefixes in specific narrow address spaces, thereby ensuring that all customer subnets can receive adequate probes, as demonstrated in Theory 1. Accordingly, to examine all the /56 customer networks of a region, as shown in Figure 7, we only need to use this quasirandom probing sequence to generate a total of 256 target /64 prefixes for probing – and no more than that.

Theory 1: For any given prefix length $x \in [48, 64]$, if the number of probes sent to a region, denoted as `offset`, satisfies the condition $\texttt{offset} < 2^{x-48}$, then each $/x$ prefix will receive at most one probe.

Proof: Let $t = x - 48$, where $t \in [0, 16]$. Assume that two different $/64$ subnets within a $/x$ prefix received

[2]A common performance metric, denoted to the ratio of discovered active addresses to sent probe packets [16], [35].

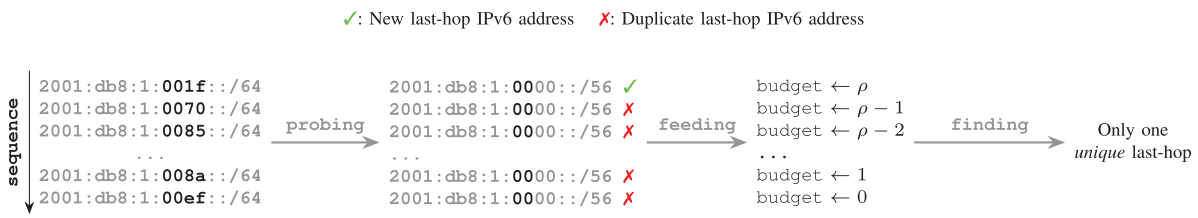


Fig. 6. An example of the inappropriate probing sequence targeting the 2001:db8:1::/48 network assigned to 256 customer subnets, each with its own /56 prefix. Note that, repeatedly probing the same customer subnet does not yield additional last-hop IPv6 router addresses.

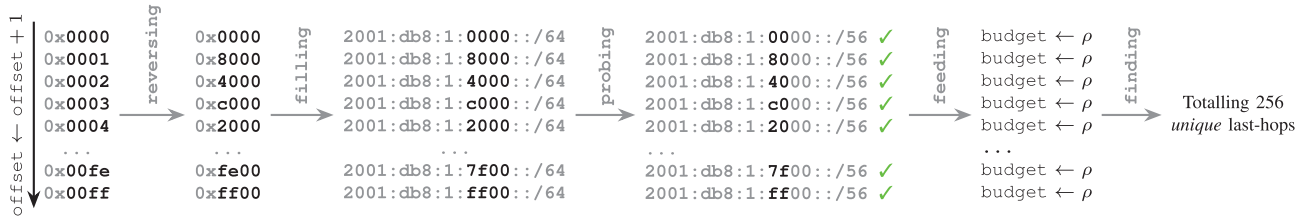


Fig. 7. *6Seeks*'s quasirandom probing sequence targeting the 2001:db8:1::/48 network assigned to 256 customer subnets, each with its own /56 prefix. Notably, each probe is dispatched to a different customer subnet.

probes numbered A and B, respectively, under the condition $\texttt{offset} < 2^t$, where $A < 2^t$ and $B < 2^t$, with $A \neq B$.

The probe numbers A and B can be expressed in binary form as:

$$A = a_{15} \cdot 2^{15} + a_{14} \cdot 2^{14} + \cdots + a_t \cdot 2^t + \cdots + a_0 \cdot 2^0$$
$$B = b_{15} \cdot 2^{15} + b_{14} \cdot 2^{14} + \cdots + b_t \cdot 2^t + \cdots + b_0 \cdot 2^0$$

where $a_i, b_i \in \{0, 1\}$ for $i \in [0, 15]$, corresponding to the bits in the range between the 48th and 64th positions of the probe targets.

The bit-reversed values of A and B are denoted as $R(A)$ and $R(B)$, and can be written as:

$$R(A) = a_0 \cdot 2^{15} + a_1 \cdot 2^{14} + \cdots + a_t \cdot 2^{15-t} + \cdots + a_{15} \cdot 2^0$$
$$R(B) = b_0 \cdot 2^{15} + b_1 \cdot 2^{14} + \cdots + b_t \cdot 2^{15-t} + \cdots + b_{15} \cdot 2^0$$

Since A and B correspond to probes within the same $/x$ prefix but belong to different $/64$ subnets, we know that A and B must share the same prefix up to the t-th bit, i.e., for $i \in [0, t-1]$, $a_i = b_i$. However, there must exist at least one bit $j \in [t, 15]$ such that $a_j \neq b_j$.

Without loss of generality, let $j \in [t, 15]$ be the first bit where $a_j = 1$ and $b_j = 0$ (or vice versa). Then, the probe number A can be written as:

$$A = a_{15} \cdot 2^{15} + \cdots + a_j \cdot 2^j + \cdots + a_t \cdot 2^t + \cdots + a_0 \cdot 2^0$$

Since $a_j = 1$, we have:

$$A \geq a_j \cdot 2^j = 2^j$$

Thus, $A \geq 2^j \geq 2^t$, because $j \geq t$. This contradicts the assumption that $A < 2^t$. Therefore, it is not possible for two probes to belong to the same $/x$ prefix if $\texttt{offset} < 2^{x-48}$. *Thus it has been demonstrated.*

In essence, our bit-reversal-based quasirandom probing sequence hypothesizes that, within an active /48 network, the address spaces that have not yet been probed are likely to correspond to *new* customer subnets. Therefore, it advocates *prioritizing searching in the least-searched directions*. To illustrate this, consider the 65,536 constituent /64 subnets of a region as points evenly distributed on a circle. Starting with point 0×0000, the next point to explore is 0×8000, as it is the least-searched compared to the others. After examining points 0×0000 and 0×8000, we proceed to sample point 0×4000, which is also the least-searched at that stage. This process continues in a similar manner until the budget is exhausted. In this way, regardless of the number of allowed probing attempts, this strategy can always direct probe packets to the subnets with the greatest scarcity of samples, thereby ensuring an even distribution of target /64 prefixes within a candidate region.

V. Evaluation & Comparison

We first conduct the real-world tests for performance evaluation of *6Seeks* and existing approaches on global IPv6 network periphery discovery, and then perform a *back-to-back* comparison between the address sets from IPv6 network periphery measurements and existing IPv6 address repositories.

A. Comparing the Existing Approaches

We use the existing approaches for IPv6 network periphery discovery, *Edgy* [14] and *Xmap* [13], as the baselines for real-world evaluation. To ensure fairness, all approaches employ ICMPv6 Echo Requests as probe packets at a consistent uplink rate of 50 kpps (kilo packets per second), and are deployed on a virtual private server which is located in Canada (AS26832) and equipped with 32 cores and 64 GB of RAM.

1) Baselines Reproduction Setup: To facilitate the performance reproduction of baselines, we supplement the following experimental setups.

- ***Edgy*.** It requires a set of /48 networks as input to effectively discover the IPv6 network periphery. However, the /48 networks used in [14] were sourced from the CAIDA

IPv6 routed /48 topology dataset[3] that was last updated on June 8, 2016. To minimize the impact of outdated data on its performance, we offered the candidate regions (the potentially-active /48 networks) derived from the first phase of *6Seeks* as *Edgy*'s input and adopted the same setups as the original study to replicate the experiments, denoted to the baseline *Edgy w/ CRs*. Additionally, *Edgy* previously used *yarrp* – a high-speed topology discovery tool [43], to perform high-speed active IPv6 topology mapping for network periphery discovery. While *yarrp* is ideal for full topology probing (i.e., identifying the sequence of router interfaces along a path to a destination), our focus is not on the entire forwarding path but only on the last hop toward an IPv6 destination. To this end, we optimized *Edgy* by sending an ICMPv6 Echo Request to a random IID within an end-user subnet, using a maximum Hop Limit of 255, in alignment with related works [7], [8]. With this optimization, *Edgy* conserved the budget that was previously wasted on the irrelevant intermediate hops, providing a solid basis for a fair performance comparison with *Xmap* and *6Seeks*. We also randomly sampled an equivalent number of /48 prefixes from global BGP prefixes as *Edgy*'s input to explore its ability to discover the IPv6 network periphery on a global scale without the aid of our candidate region collection, denoted to the baseline *Edgy w/o CRs*.

- ***Xmap*.** This technique involves arbitrarily sampling the random-IID targets within specified prefixes to elicit *indirect responses* from last-hop IPv6 routers (IPv6 periphery devices). However, its IPv6 prefixes used in [13] for probing have not yet been published, posing significant challenges for reproducing the experiments. To address this limitation, we leverage the *6Seeks* candidate regions as input for *Xmap*, denoted to the baseline *Xmap w/ CRs*. Additionally, to evaluate *Xmap*'s performance in the absence of *6Seeks* candidate regions, we randomly sample an equivalent number of /48 prefixes from global BGP prefixes as an alternative input, denoted to the baseline *Xmap w/o CRs*.

2) Parameter Optimization: In the 6Seeks dynamic region probing algorithm, the parameter ρ adjusts the probe limit for candidate /48 prefixes. To optimize ρ, we sampled one million /48 prefixes from the candidate set and evaluated 6Seeks performance for $\rho \in \{2, 4, 8, 16, 32, 64, 128\}$. Figure 9 illustrates the number of unique last-hop IPv6 addresses for each ρ. Clearly, $\rho = 64$ lies at an inflection point, achieving comparable address discovery to $\rho = 128$ using half the probes. Thus, we set $\rho = 64$ as the *6Seeks* budget pool capacity.

3) Overall Performance Comparison: In October 2024, we conducted comprehensive real-world measurements to evaluate the performance of all the baselines and *6Seeks*. Note that we would not impose the explicit budget limitation on the measurement campaigns, as prematurely stopping the process can lead to incomplete results and skewed performance evaluations. For example, slower-converging methods may seem less effective due to incomplete measurements that don't capture their true performance. To this end, we use the *efficiency*, defined as the ratio of unique last-hop router (IPv6 periphery device) addresses discovered to the total number of probes sent, for the metric of performance comparison.

[3]https://www.caida.org/catalog/datasets/ipv6_routed_48_topology_dataset/

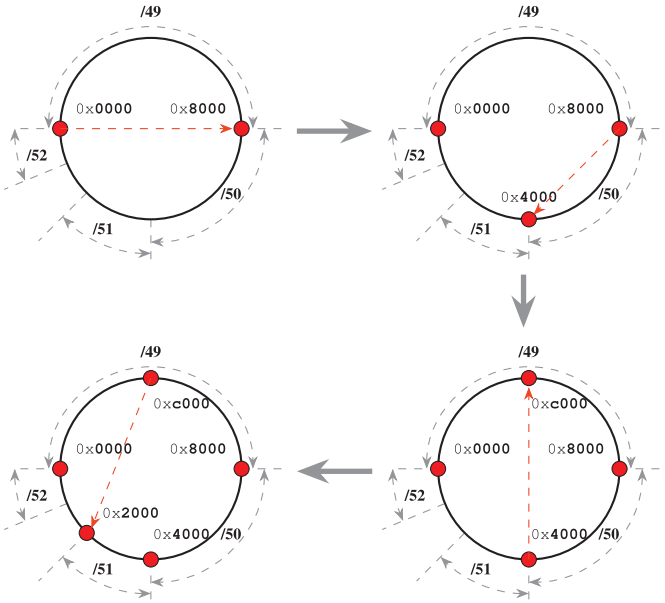


Fig. 8. An illustration of navigating the address space of a /48 network through our quasirandom probing sequence.

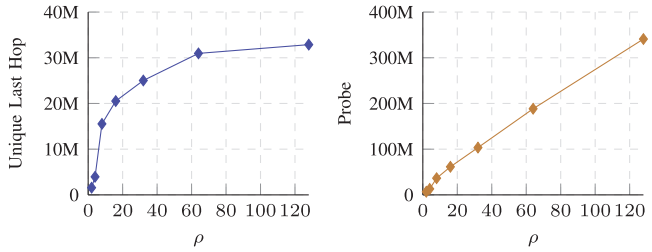


Fig. 9. The number of unique last-hop IPv6 addresses and probes for the selected budget $\rho \in \{2, 4, 8, 16, 32, 64, 128\}$.

TABLE II
THE RESULTS OF EDGY (BASELINE) PER ROUND ON CANDIDATE REGIONS

Round	$\eta^{\dagger}$	Region	Budget	Involved /48*	Last Hop*
/56	16	4,473,239	1.14B	2,969,639	28,888,569
/60	16	226,391	0.93B	236,299	55,932,296
/62	4	32,798	0.54B	131,468	43,010,455
/64	-	17,862	1.17B	213,585	85,481,254

†: The parameter remains consistent with [14].
*: Results from different rounds may overlap.

First, we conducted the IPv6 network periphery discovery using *Edgy* and *Xmap* on randomly-sampled /48 IPv6 networks to evaluate their performance when applied individually to the global-scale measurements. As shown in Table III, both methods suffer from inefficiency, discovering a maximum of 198,500 last-hop IPv6 router addresses – several orders of magnitude less than other baselines. In other words, straightforwardly using existing methods for global IPv6 network periphery discovery may not be effective. Therefore, we excluded these results from the subsequent address analysis due to their limited number of addresses.

TABLE III
OVERALL PERFORMANCE COMPARISON OF 6SEEKS AND BASELINES

Approach	Probe	Time	Last Hop	Efficiency
Xmap w/o CRs	1.15B	6:23:48	37.1k	0.003%
Edgy w/o CRs	1.15B	6:23:48	198.5k	0.017%
Xmap w/ CRs	3.78B	21:00:08	35.6M	0.94%
Edgy w/ CRs	3.78B	21:00:08	**128.0M**	**3.38%**
Xmap w/ CRs	1.40B	7:46:40	22.9M	1.64%
6Seeks	1.40B	7:46:40	**128.3M**	**9.16%**

Notes: CRs = 6Seeks candidate regions.

Next, we take the candidate regions, i.e., the 4,473,239 potentially-active /48 networks identified in § III, as input to evaluate the performance of *6Seeks* and the baseline methods in the global IPv6 network periphery discovery. For fairness, we replicated *Edgy*'s experiments following the methodology outlined in [14]: The candidate /48 networks were systematically probed at increasingly finer granularities – /56, /60, /62, and /64 – until a predefined stopping condition was satisfied. Specifically, if the number of unique last-hop router addresses identified during a probing round exceeded the threshold η, the responsive prefixes were further subdivided for additional probing in the subsequent round. The results of the baseline *Edgy w/ CRs* for each round are detailed in Table II. After deduplicating across rounds, *Edgy* successfully identified approximately 128.0 million unique IPv6 router addresses, using a total of 3.78 billion probes. As a comparison, *6Seeks* consumed approximately 1.4 billion probe packets and discovered 128.3 million unique last-hop IPv6 router addresses.

Last, we evaluated the *Xmap*'s performance using the same probe budgets as *6Seeks* and *Edgy*, respectively. The results demonstrate that, with the aid of candidate regions provided by *6Seeks*, *Xmap* effectively identified approximately 22.9 million and 35.6 million unique last-hop IPv6 router addresses, respectively, while sending a total of 1.40 billion and 3.78 billion probes.

In summary, *6Seeks* outperforms all the baselines in terms of the *efficiency* of global IPv6 network periphery measurement, and can identify over 128 million unique last-hop IPv6 router addresses using only 37% of the probing resources required by the state-of-the-art approach (*Edgy*).

4) Convergence Speed: Figure 10 presents the cumulative growth in the number of unique last-hop IPv6 routers discovered over budget consumption. The results clearly demonstrate that, in comparison to the baseline methods, *6Seeks* consistently achieves the highest discovery efficiency throughout the entire measurement campaign. This indicates that *6Seeks* not only outperforms all baselines in global-scale measurement tasks but also delivers optimal results when the measurement scale is constrained. Therefore, we recommend *6Seeks* for IPv6 network periphery discovery, as it effectively minimizes both time and probe resource consumption.

5) Subnet Inference: In this work, we rely heavily on the notion of address discriminating prefix length (DPL)[4] for inferring the customer subnet like [43]. The Discriminating

[4]Kohler et al. [44] introduced the term "discriminating prefix length." It has been employed widely in the structure analysis of both active and passive measurements.

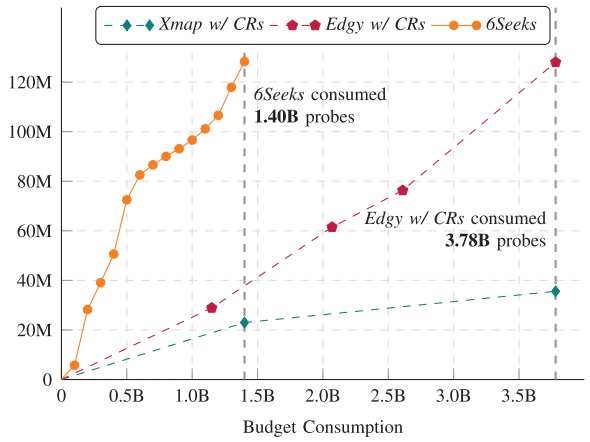


Fig. 10. The cumulative number of unique last-hop IPv6 addresses discovered by *6Seeks* and baselines. *6Seeks* only used 1.4 billion probes to accomplish the global IPv6 network periphery discovery, while *Edgy* required 3.78 billion.

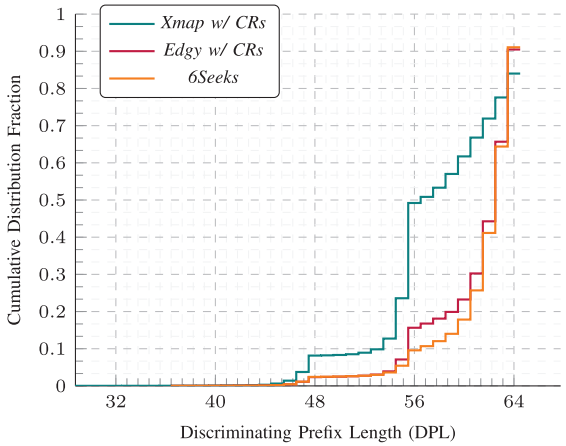


Fig. 11. Discriminating Prefix Length (DPL) Distributions for the address sets of *6Seeks* and baselines.

Prefix Length (DPL) of an IPv6 address is the position of the first bit (starting from the leftmost, or most significant) where it differs from its nearest neighbor in a sorted set, e.g., the last-hop IPv6 router addresses after sorting. This represents how many bits must be compared, from left to right, to distinguish one address from another. For example, a primitive router might use the DPL to determine how to forward traffic when two addresses require different routing decisions. For addresses in different subnets, their DPL essentially reflects the proximity of their subnets: higher DPLs indicate closer subnets. As the IPv6 periphery addresses typically belong to different customer subnets, DPL provides a lower bound for the prefix length of their respective subnets.

Figure 11 presents the prefix length distribution of customer subnets inferred from DPLs. It is evident that most customer subnets fall within the range of /48 to /64. Notably, over 90% subnets whose respective addresses are identified by *6Seeks* and *Edgy* have prefix lengths longer than /56, which are typically deployed at the end-sites where customers are located rather than being used within the backbone Internet. This shared pattern suggests that providers prefer smaller customer subnet allocations to enhance the utilization of IPv6 address

TABLE IV
THE RESULTS OF BACK SCANNING AND NEIGHBOR SCANNING ON ADDRESS SETS OF 6SEEKS AND BASELINES

Source	Back Scanning			Neighbor Scanning		
	Hit (IID Entropy)	*Miss (IID Entropy)*	Churn Rate	Total	*New (IID Entropy)*	Retention
Xmap w/ CRs	13.0M (0.5046)	22.6M (0.7003)	63.5%	55.6M	53.3M (0.7443)	**2.358**
Edgy w/ CRs	26.7M (0.5831)	101.3M (0.7064)	79.1%	107.6M	94.7M (0.7235)	**0.935**
6Seeks	32.2M (0.6098)	96.1M (0.7146)	74.9%	107.3M	91.1M (0.7257)	**0.947**

Notes: Churn Rate = $\frac{Miss}{Hit + Miss}$. Retention (Ratio) = $\frac{New}{Miss}$. *CRs* = *6Seeks* Candidate Regions.

space. However, *Xmap*'s findings deviate from this trend. This discrepancy arises because its randomized probing strategy makes it easier to dispatch probes to the IPv6 periphery devices associated with large customer subnets. For example, within a /48 network, the probability of hitting a /64 subnet is $\frac{1}{256}$ compared to that for a /56 subnet. As a result, *Xmap* discovered far fewer last-hop router addresses than *6Seeks* and *Edgy*, resulting in its findings being disproportionately concentrated on larger subnets (e.g., over 50% of *Xmap*'s subnets inferred from DPLs have prefix lengths of /56 or shorter) which is deviating from the truth.

6) Address Churn & Retention: To protect customers' privacy from IP tracking attacks, ISPs widely adopt prefix rotation in IPv6 networks [3], [45], causing the addresses of IPv6 periphery devices to change frequently – a phenomenon called address churn [46]. To investigate the impact of prefix rotation on IPv6 studies, we intentionally delayed the address churn analysis by 30 days following our global measurement campaigns to ensure complete lost of those volatile addresses.

Back Scanning. When the 30-day interval concluded, we sent ICMPv6 Echo Requests to the last-hop IPv6 router addresses and then received the *direct responses* (ICMPv6 Echo Replies from the probed targets), so as to analyze how many IPv6 periphery addresses remain active, denoted as *Hit*, and how many are lost, denoted as *Miss*. Our findings indicate that, on average, 75.3% IPv6 periphery addresses became inactive after a 30-day interval, underscoring a substantial rate of address churn among last-hop IPv6 routers. This trend highlights the effectiveness of widespread prefix rotation mechanisms in protecting customer privacy, which is a positive outcome from a security and privacy perspective.

Neighbor Scanning. While address churn offers certain advantages, it also presents a challenge for studying the IPv6 network periphery. A potential concern is that outdated data may lose value as most addresses become unresponsive within 30 days. To address this, we proposed to conduct the neighbor scanning. Specifically, we hypothesize that, for a given device, the IP addresses assigned before and after a prefix rotation typically belong to the same customer subnet, unless the device's physical location has changed. Although the previous address becomes inactive, scanning within the customer subnet (i.e., the neighboring address space) is highly likely to reveal the new address. Since most customer subnets are /56 or smaller, as shown in § V-A5, we consolidate these *Miss* addresses into a set of /56 prefixes by truncating them to the 56 most significant bits and removing duplicates. Subsequently, we exhaustively probe all potential /64 subnets within these /56 prefixes by appending random Interface Identifiers (IIDs), and receivce the indirect responses from last-hop routers to obtain their new addresses. Using this approach, we identified over 107 million last-hop router addresses from neighbor scans on *Edgy*'s and *6Seeks*'s *Miss* addresses, with more than 91 million being previously unseen (*New*), as shown in Table IV. This demonstrates that simply scanning the neighboring /56 address space can efficiently recover an average of 94% of volatile addresses, even after the 30-days interval; *Xmap* appears a significantly higher retention ratio due to the much smaller total number of its prior addresses compared to other methods. In other words, while the widespread prefix rotation make a significant portion of the IPv6 periphery address becoming inactive quickly, researchers can easily recover the equivalent number of IPv6 addresses through neighbor scanning on the churn ones.

The configuration of an IPv6 address can affect both its lifespan and addressing pattern [47], [48]. For example, IPv6 addresses using port-embedded IIDs are typically used by organizations scanning the internet to highlight their benign intentions [49], thus being regarded as permanent addresses and not expiring. Conversely, IPv6 addresses using randomized IIDs are commonly used for modern mobile devices [3], [45], [50] that frequently change addresses to protect privacy, thus having a shorter lifespan. In other words, the IID of an address might reflect its lifespan. To illustrate this, we analyzed the address sets from all three sources (*Xmap*, *Edgy*, and *6Seeks*).

Figure 12 displays the IID entropy distribution of the IPv6 periphery addresses, broken down by whether the address was responsive to back scanning (*Hit*) or not (*Miss*), or newly discovered from neighbor scanning (*New*). In the addresses from these three sources, all *Hit* address sets had lower IID entropy than their counterpart *Miss* ones, suggesting that higher IID entropy makes an address less likely to be permanent. The *New* addresses had a similar IID entropy distribution to the *Miss* addresses. Although we cannot directly match newly discovered addresses to lost ones due to the absence of device identifiers, their similar patterns suggest that they could belong to the same group of IPv6 periphery devices. *In a nutshell*, a high-entropy interface identifier typically suggests a short lifespan for the IPv6 address, or vice versa.

B. Comparing the Existing IPv6 Address Repositories

The results from global IPv6 network periphery measurements have not only expanded the IPv6 address corpus but have also significantly provided novel IPv6 insights. To showcase these advancements, we meticulously compared

TABLE V
OVERALL COMPARISON OF THE IPV6 ADDRESS SETS

Source	IPv6 Address							Autonomous System						
	Involved /64	IID Entropy	EUI-64		Aliased		Total	Top 1		Top 2		>1k IPs	Total
IPv6 Hitlist	4.9M	0.3250	697.3k	7.5%	4.8k	0.5%	9.3M	17.5%	♦	5.0%	▲	525	19177
AddrMiner	4.2M	0.6796	946.1k	1.3%	54.9M	73.9%	74.3M	88.3%	▼	2.2%	♦	523	21905
Hmap	40.4M	0.2350	355.7k	0.3%	9.4M	8.7%	107.1M	13.2%	★	8.2%	●	645	16901
Xmap w/ CRs	29.9M	0.6289	18.0M	50.6%	0.5k	-	35.6M	15.8%	♦	8.9%	▲	609	6996
Edgy w/ CRs	115.8M	0.6807	58.6M	45.8%	48.5k	-	128.0M	20.4%	▼	6.9%	★	761	8497
6Seeks	116.4M	0.6870	61.4M	47.7%	7.1k	-	128.3M	22.9%	▼	6.6%	●	768	8439

♦ Free SAS (AS12322), ▲ V. tal (AS7738), ▼ Amazon (AS16509), ★ Developed Methods LLC (AS400519), ● Digital Workspaces (AS397165)
♦ Orange S.A. (AS3215), ▲ Sky Broadband (AS5607), ▼ China Mobile (AS9808), ★ Reliance Jio (AS55836), ● China Unicom (AS4837)

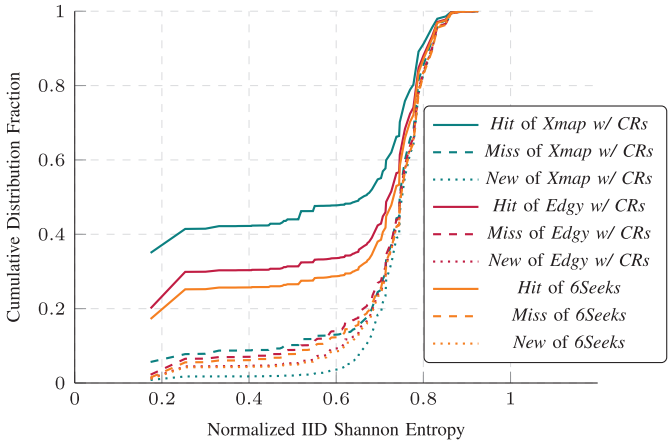


Fig. 12. The IID entropies of IPv6 address sets of *6Seeks* and baselines, divided by their status in the back scanning and neighbour scanning. The IPv6 periphery addresses that remain active after a 30-day interval, i.e., *Hits*, generally exhibit lower IID entropies.

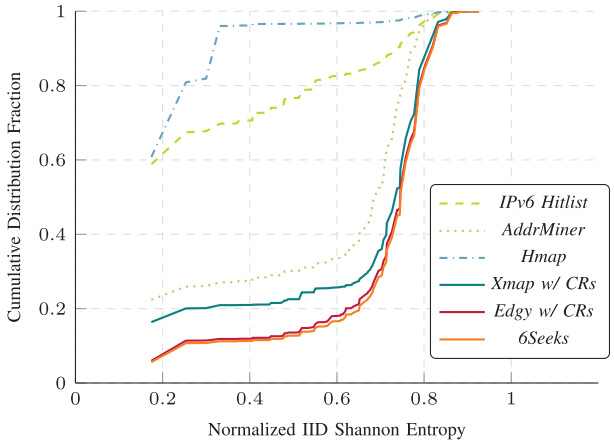


Fig. 14. The IID entropy distribution of the IPv6 addresses from different datasets. The address sets from IPv6 network periphery measurements have higher entropies on average than the existing IPv6 address repositories (except for *AddrMiner*).

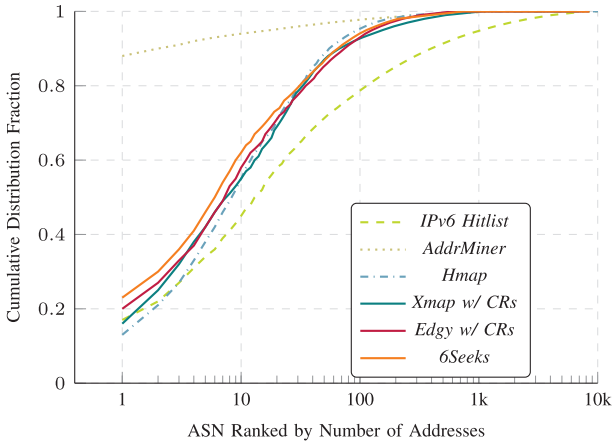


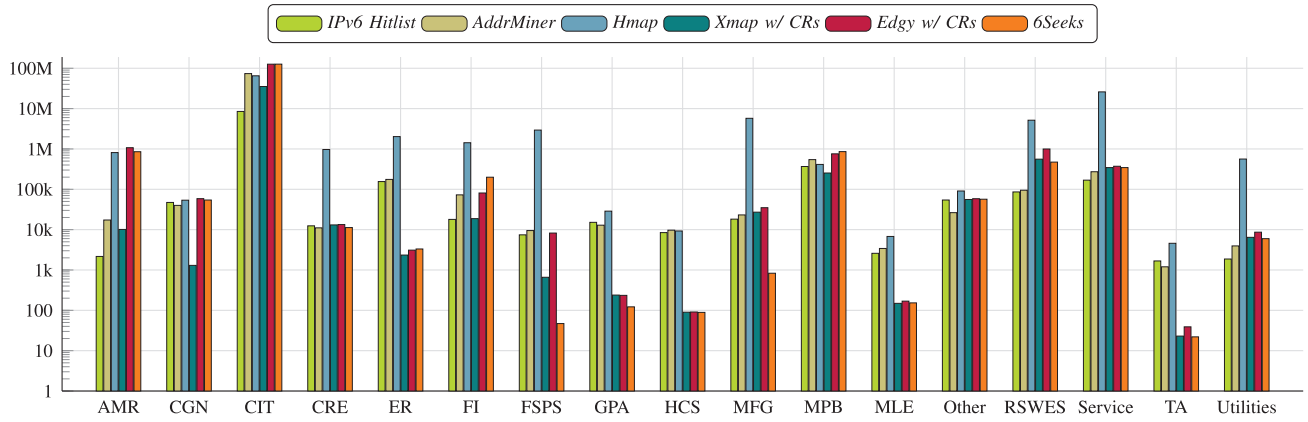
Fig. 13. Distribution of addresses from IPv6 network periphery measurements and existing repositories across autonomous systems (ASes). Note that the majority of *AddrMiner* addresses are attributed to one single AS.

findings from global IPv6 network periphery measurements on the following metrics with existing publicly accessible IPv6 address repositories, including the address set of *Hmap* [18], the address set of *AddrMiner* [19], [23], and the *IPv6 Hitlist* Service data [22].

1) Address Space Coverage: Understanding the distribution of IPv6 addresses is a key objective in IPv6 measurement studies. The value of a dataset increases with the extent of its covered address space, as it provides critical information about IPv6 address allocation. Following established conventions [18], [21], we employ the number of individual /64 prefixes as the metric to evaluate the IPv6 address space coverage of *6Seeks*'s IPv6 address set and other datasets. As shown in Table V, the IPv6 network periphery measurements by *Edgy* and *6Seeks* achieve the largest address coverage compared to other datasets, involving 115.8 million and 116.4 million /64 prefixes, respectively – representing at least a 186% improvement over the state-of-the-art dataset. Therefore, *6Seeks*'s IPv6 address set surpasses all existing IPv6 address repositories in terms of address space coverage.

Figure 16 presents a joint overlap analysis of /64 address space coverage across the datasets. There is almost no overlap between the /64 prefixes identified by IPv6 network periphery measurements (e.g., *Xmap*, *Edgy*, and *6Seeks*) and those in existing IPv6 address repositories. Notably, the /64 prefixes discovered exclusively by *6Seeks*, exclusively by *Edgy*, or jointly by both contribute to the majority of the covered address space. Thus, conducting IPv6 network periphery scanning can effectively explore IPv6 address spaces beyond the reach of traditional IPv6 scanning methods, thereby enhancing the comprehensiveness of IPv6 active measurement.

2) As-Level Distribution: The involved Autonomous Systems (ASes) play a key role in assessing the quality of IPv6 address sets. However, relying solely on the raw count of ASes



Notes: AMR = Agriculture, Mining, and Refineries. CGN = Community Groups and Nonprofits. CIT = Computer and Information Technology. CRE = Construction and Real Estate. ER = Education and Research. FI = Finance and Insurance. FSPS = Freight, Shipment, and Postal Services. GPA = Government and Public Administration. HCS = Health Care Services. MFG = Manufacturing. MPB = Media, Publishing, and Broadcasting. MLE = Museums, Libraries, and Entertainment. RSWES = Retail Stores, Wholesale, and E-commerce Sites. TA = Travel and Accommodation.

Fig. 15. The counts of IPv6 addresses from different datasets by ASdb industry categories.

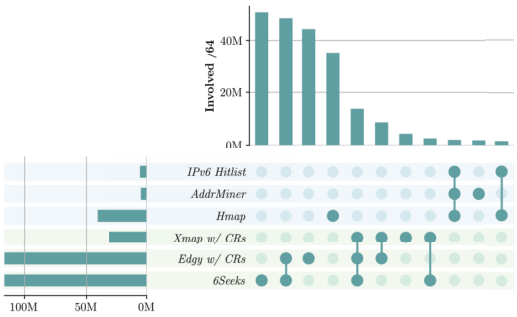


Fig. 16. The joint overlap analysis of /64 prefixes involved by the address sets from IPv6 network periphery measurements and existing repositories.

can lead to inflated performance metrics due to the highly uneven AS-level distribution of IPv6 addresses. For instance, the *AddrMiner* dataset claims to include 21,904 ASes, yet 88.3% of its IPv6 addresses come from a single AS, while 9,674 ASes contribute fewer than five addresses each. To mitigate this potential distortion in quantitative comparison, we additionally consider the number of ASes with over 1,000 addresses. Results indicate that the address sets of *6Seeks* and *Edgy* encompass 758 and 761 ASes, respectively, each having more than 1,000 addresses, while the state-of-the-art dataset *Hmap* has only 645 ASes. Figure 13 illustrates the cumulative distribution fraction of all addresses in each dataset across different Autonomous Systems (ASes). Most datasets exhibit an even AS-level distribution of address quantity, similar to each other. However, the *AddrMiner* dataset is significantly top-heavy, with majority of its addresses concentrated in the top-1 AS, setting it apart from other datasets.

Table V presents the top two autonomous systems (ASes) for each dataset. It is evident that address sets from different sources are typically associated with distinct top ASes. For instance, IPv6 network periphery measurements predominantly identify addresses concentrated in ASes belonging to customer ISPs, such as China Mobile (AS9808), Reliance Jio (AS55836), and China Unicom (AS4837). In contrast, existing IPv6 repositories tend to favor ASes associated with hosting data centers and cloud services, such as Amazon (AS16509) and Developed Methods LLC (AS400519).

To explore this discrepancy thoroughly, we examine the industry categories of ASes from which all the addresses originate, as classified by ASdb [51], and accumulate the addresses per dataset by these categories. As shown in Figure 15, while the top AS category of addresses is consistent across all datasets, i.e., "Computer and Information Technology" (CIT), the discrepancy appears in the AS sub-categories of CIT addresses. As shown in Table VI, 34.49% and 38.29% of the addresses from *Edgy*'s and *6Seeks*'s discoveries originate from the "Phone Provider" sub-category. In contrast, only 3.69% of the addresses in the *IPv6 Hitlist* originate from "Phone Provider" ASes, and even fewer for the other two existing datasets (*AddrMiner* and *Hmap*). However, 16.46% of the *IPv6 Hitlist* addresses, 89.43% of the *AddrMiner* addresses, and 32.87% of the *Hmap* addresses belong to "Hosting and Cloud Provider" ASes, while the IPv6 periphery addresses rarely originate from such ASes. This discrepancy is likely due to the nature of the three existing IPv6 address datasets, which are usually collected from multiple sources, such as target generation, DNS records, network topology, and passive sources. Our data, coming from the last-hop IPv6 routers, is concentrated in the autonomous systems where the IPv6 periphery devices exist, typically in customer ISPs. As such, our IPv6 periphery address sets have a significantly different AS-level distribution compared to existing IPv6 address repositories.

3) IID Analysis: IPv6 addresses using EUI-64 Interface Identifiers (IIDs) are directly derived from a device's hardware MAC address [29], which exposes the device's Layer-2 information at Layer-3. To isolate these EUI-64 addresses from others, we check if the fourth and fifth bytes of the IID are `0xFF` and `0xFE`, respectively. The probability that a randomly-generated IID will match these bytes is 2^{-16}. As such, we would expect $\frac{61,400,000}{65,536}$ addresses (<1,000) using randomly-generated IIDs but matching this pattern even in the *6Seeks*'s address set. Table V presents the number of

TABLE VI
RATIO OF IPV6 ADDRESSES BY ASDB SUB-CATEGORIES IN THE CIT AUTONOMOUS SYSTEMS

Source	CNS	HCP	ISP	Other	PP	SC	Search	SD	TCS
IPv6 Hitlist	0.29%	**16.46%**	**70.14%**	3.50%	3.69%	0.01%	0.94%	4.83%	0.15%
AddrMiner	0.89%	**89.43%**	**7.32%**	0.41%	0.51%	0.00%	0.12%	1.27%	0.06%
Hmap	0.39%	**32.87%**	**54.48%**	0.01%	0.58%	0.00%	0.13%	11.48%	0.08%
Xmap w/ CRs	0.00%	2.39%	**84.87%**	0.00%	**12.57%**	0.01%	0.01%	0.11%	0.04%
Edgy w/ CRs	0.00%	1.29%	**63.38%**	0.00%	**34.49%**	0.00%	0.01%	0.03%	0.81%
6Seeks	0.00%	1.02%	**59.95%**	0.00%	**38.29%**	0.00%	0.01%	0.05%	0.68%

Notes: CNS = Computer and Network Security. HCP = Hosting and Cloud Provider. ISP = Internet Service Provider. PP = Phone Provider. SC = Satellite Communication. SD = Software Development. TCS = Technology Consulting Services. CIT = Computer and Internet Technology.

EUI-64 IPv6 addresses isolated from all address datasets. Results show that, none of the existing IPv6 address repositories (i.e., *IPv6 Hitlist*, *AddrMiner*, and *Hmap*) can provide over one million EUI-64 IPv6 addresses, while *6Seeks*, *Edgy*, and *Xmap* contribute 61.4 million, 58.6 million, and 18.0 million EUI-64 IPv6 addresses, respectively. This demonstrates that a significant portion of IPv6 periphery devices (at least 45.8%) worldwide still use legacy EUI-64 addressing, which is vulnerable to device tracking attacks.

Figure 14 plots the cumulative distribution function (CDF) of all addresses found in each dataset over IID entropy. It is immediately apparent that the corpus from IPv6 network periphery measurements exhibits significantly higher IID entropy compared to the *IPv6 Hitlist* and *Hmap*, primarily due to their high ratio of EUI-64 addresses. The *AddrMiner* dataset also exhibits a high IID entropy but it significantly lacks EUI-64 addresses, indicating that most of its addresses use the randomized IIDs. In summary, the results of our global IPv6 network periphery measurements provide the community with an IPv6 address set whose IID distribution is different from that of existing IPv6 address repositories, thereby significantly enriching our IPv6 corpus.

4) Aliased Ratio: It is crucial to promptly eliminate aliased addresses in IPv6 measurements. Probing within an IPv6 aliased prefix[5] generates disproportionately high responses, appearing as though they come from a large number of devices when, in fact, they originate from a single host. This significantly distorts the results from the ground truth, and we do not want IPv6 aliased addresses to appear in the address sets. Using the aliased and non-aliased prefixes provided by Zirngibl et al. [22] and Gasser et al. [28], we identified aliased addresses within existing IPv6 address repositories and the address sets from global IPv6 network periphery measurements. The results indicate that the ratio of aliased addresses is negligible in *6Seeks*'s, *Edgy*'s, and *Xmap*'s address sets. In contrast, the *IPv6 Hitlist* and *Hmap* datasets have aliased ratios of 0.5% and 8.7%, respectively. An exception is the *AddrMiner* dataset, where 73.9% of its addresses are detected as aliased, with most originating from the AS16509 (Amazon). This demonstrates that *AddrMiner* lacks resistance to IPv6 aliased prefixes when applied to IPv6 scanning. Therefore, compared to existing IPv6 address repositories, our IPv6 periphery address sets have a far lower aliased ratio.

To conclude, the findings of *6Seeks* system not only outperform existing IPv6 address repositories in quantity but also offer numerous distinctive characteristics, significantly complementing existing IPv6 corpora.

VI. ETHICAL CONSIDERATIONS AND DISCUSSIONS

We adhere to the principles of responsible Internet conduct, as recommended by Partridge and Allman [52]. To minimize the impact on target networks, we implement the following measures:

- The probing uplink bandwidth is limited to 50 kpps, set through negotiations with VPS providers to ensure unaffected local network operations.
- Each /64 prefix is probed once in a quasirandom permutation to disperse scanning traffic across the address space, minimizing impact on target networks.
- A website URL is embedded in probe packet payloads to address abuse complaints and exclude entities upon request. Privacy-conscious network administrators can contact us via the email address listed on the site.
- Active /48 prefixes and *6Seeks* source code will be released, but global IPv6 periphery device addresses, lacking dedicated security, will not be disclosed to prevent cybercriminal abuse.

Internet censorship typically contributes substantial biases to active network measurements. However, as ICMPv6 Echo Requests used by *6Seeks* are less intrusive, IPv6 periphery measurements are minimally affected by censorship mechanisms like the GFW. For instance, when probing from outside China, China Mobile (AS9808) and China Unicom (AS4837) – two major Chinese Internet service providers – dominate, accounting for about 30% of the IPv6 periphery addresses discovered by *6Seeks*. This indicates that the Internet censorship does not significantly block our probes from the Canada VPS.

VII. CONCLUSION

This paper presents *6Seeks*, an IPv6 scanning system designed for fast IPv6 network periphery discovery on a global scale. Without requiring seed IPv6 addresses, it employs a heuristic method to collect active /48 networks from global BGP prefixes. Subsequently, it adopts a reinforcement learning-based dynamic probing strategy to automatically

[5]An IPv6 prefix, under which every IP address replies to queries, commonly used in CDNs and enterprise networks [28].

adjust the allocation of probe resources across these networks. Real-world tests demonstrate that *6Seeks* outperforms all baseline solutions. Using only 37% of the probing resources required by the current state-of-the-art solution, *6Seeks* successfully identified more than 128 million last-hop IPv6 router addresses. Compared to existing public datasets, the IPv6 addresses identified by *6Seeks* are more numerous and exhibit many unique characteristics, significantly enriching our IPv6 corpus.

ACKNOWLEDGMENT

The authors would like to greatly appreciate the anonymous reviewers for their insightful comments.

REFERENCES

[1] M. Holdrege and P. Srisuresh, *IP Network Address Translator (NAT) Terminology and Considerations*, document RFC 2663, Aug. 1999. [Online]. Available: https://www.rfc-editor.org/info/rfc2663

[2] R. Hinden and S. Deering, *IP Version 6 Addressing Architecture*, document RFC 4291, Feb. 2006. [Online]. Available: http://www.rfc-editor.org/rfc/rfc4291.txt

[3] C. Byrne, D. Drown, and A. Vizdal, *Extending an IPv6 /64 Prefix From a Third Generation Partnership Project (3GPP) Mobile Interface to a LAN Link*, document RFC 7278, Jun. 2014.

[4] E. Davies and J. Mohacsi, *Recommendations for Filtering ICMPv6 Messages in Firewalls*, document RFC 4890, May 2007. [Online]. Available: http://www.rfc-editor.org/rfc/rfc4890.txt

[5] H. Singh, W. Beebee, C. Donley, and B. Stark, *Basic Requirements for IPv6 Customer Edge Routers*, document RFC 7084, Nov. 2013.

[6] D. Plonka and A. Berger, "Temporal and spatial classification of active IPv6 addresses," in *Proc. Internet Meas. Conf.*, Oct. 2015, pp. 509–522.

[7] E. Rye, R. Beverly, and K. C. Claffy, "Follow the scent: Defeating IPv6 prefix rotation privacy," in *Proc. 21st ACM Internet Meas. Conf.*, Nov. 2021, pp. 739–752.

[8] E. C. Rye and R. Beverly, "IPvSeeYou: Exploiting leaked identifiers in IPv6 for street-level geolocation," in *Proc. IEEE Symp. Secur. Privacy (SP)*, May 2023, pp. 3129–3145.

[9] J. Czyz, M. Luckie, M. Allman, and M. Bailey, "Don't forget to lock the back door! A characterization of IPv6 network security policy," in *Proc. Netw. Distrib. Syst. Secur. Symp.*, Feb. 2016.

[10] F. Gont and T. Chown, *Network Reconnaissance in IPv6 Networks*, document RFC 7707, Mar. 2016.

[11] L. Pan et al., "Your router is my prober: Measuring IPv6 networks via ICMP rate limiting side channels," in *Proc. Netw. Distrib. Syst. Secur. Symp.*, Feb. 2023.

[12] A. Conta, S. Deering, and M. Gupta, *Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification*, document RFC 4443, Mar. 2006. [Online]. Available: http://www.rfc-editor.org/rfc/rfc4443.txt

[13] X. Li, B. Liu, X. Zheng, H. Duan, Q. Li, and Y. Huang, "Fast IPv6 network periphery discovery and security implications," in *Proc. 51st Annu. IEEE/IFIP Int. Conf. Dependable Syst. Netw. (DSN)*, Jun. 2021, pp. 88–100.

[14] E. C. Rye and R. Beverly, "Discovering the IPv6 network periphery," in *Proc. PAM*, Jan. 2020, pp. 3–18.

[15] T. Narten, G. Huston, and L. Roberts, *IPv6 Address Assignment to End Sites*, document BCP 157, Mar. 2011.

[16] 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 *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, May 2021, pp. 1–10.

[17] 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 Trans. Netw.*, vol. 31, no. 4, pp. 1870–1885, Apr. 2023.

[18] B. Hou, Z. Cai, K. Wu, T. Yang, and T. Zhou, "Search in the expanse: Towards active and global IPv6 hitlists," in *Proc. IEEE Conf. Comput. Commun. (IEEE INFOCOM)*, May 2023, pp. 1–10.

[19] G. Song et al., "AddrMiner: A comprehensive global active IPv6 address discovery system," in *Proc. USENIX ATC*, 2022, pp. 309–326.

[20] T. Yang and Z. Cai, "Efficient IPv6 router interface discovery," in *Proc. IEEE Conf. Comput. Commun. (IEEE INFOCOM)*, May 2024, pp. 1641–1650.

[21] T. Yang, B. Hou, Y. Yang, and Z. Cai, "Sweeping the IPv6 Internet: High-efficiency router interface discovery with weighted sampling," *IEEE Trans. Netw.*, vol. 33, no. 1, pp. 271–285, Feb. 2025.

[22] J. Zirngibl, L. Steger, P. Sattler, O. Gasser, and G. Carle, "Rusty clusters: Dusting an IPv6 research foundation," in *Proc. 22nd ACM Internet Meas. Conf.*, Oct. 2022, pp. 395–409.

[23] G. Song et al., "AddrMiner: A fast, efficient, and comprehensive global active IPv6 address detection system," *IEEE/ACM Trans. Netw.*, vol. 32, no. 5, pp. 3870–3887, Oct. 2024.

[24] M. Stubbig, *Looking Glass Command Set*, document RFC 8522, Feb. 2019. [Online]. Available: https://www.rfc-editor.org/info/rfc8522

[25] E. Jasinska, N. Hilliard, R. Raszuk, and N. Bakker, *Internet Exchange BGP Route Server*, document RFC 7947, Sep. 2016. [Online]. Available: https://www.rfc-editor.org/info/rfc7947

[26] Y. Rekhter, S. Hares, and T. Li, *A Border Gateway Protocol 4 (BGP-4)*, document RFC 4271, Jan. 2006. [Online]. Available: https://www.rfc-editor.org/info/rfc4271

[27] (2024). *University of Oregon Route Views Project*. Accessed: Oct. 2024. [Online]. Available: http://www.routeviews.org/

[28] O. Gasser et al., "Clusters in the expanse: Understanding and unbiasing IPv6 hitlists," in *Proc. Internet Meas. Conf.*, Oct. 2018, pp. 364–378.

[29] S. Thomson, T. Narten, and T. Jinmei, *IPv6 Stateless Address Autoconfiguration*, document RFC 4862, Sep. 2007. [Online]. Available: http://www.rfc-editor.org/rfc/rfc4862.txt

[30] P. Foremski, D. Plonka, and A. Berger, "Entropy/IP: Uncovering structure in IPv6 addresses," in *Proc. Internet Meas. Conf.*, Nov. 2016, pp. 167–181.

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

[32] 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," *Comput. Netw.*, vol. 203, Feb. 2022, Art. no. 108666.

[33] T. Yang, Z. Cai, B. Hou, and T. Zhou, "6Forest: An ensemble learning-based approach to target generation for Internet-wide IPv6 scanning," in *Proc. IEEE Conf. Comput. Commun. (IEEE INFOCOM)*, May 2022, pp. 1679–1688.

[34] A. Murdock, F. Li, P. Bramsen, Z. Durumeric, and V. Paxson, "Target generation for Internet-wide IPv6 scanning," in *Proc. Internet Meas. Conf.*, Nov. 2017, pp. 242–253.

[35] G. Song et al., "DET: Enabling efficient probing of IPv6 active addresses," *IEEE/ACM Trans. Netw.*, vol. 30, no. 4, pp. 1629–1643, Aug. 2022.

[36] G. Williams et al., "6Sense: Internet-wide IPv6 scanning and its security applications," in *Proc. USENIX Secur.*, Aug. 2024, pp. 2281–2298.

[37] G. Williams and P. Pearce, "Seeds of scanning: Exploring the effects of datasets, methods, and metrics on IPv6 Internet scanning," in *Proc. ACM Internet Meas. Conf.*, Nov. 2024, pp. 295–313.

[38] S. J. Saidi, O. Gasser, and G. Smaragdakis, "One bad apple can spoil your IPv6 privacy," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 52, no. 2, pp. 10–19, Apr. 2022.

[39] E. Rye and D. Levin, "IPv6 hitlists at scale: Be careful what you wish for," in *Proc. ACM SIGCOMM Conf.*, Sep. 2023, pp. 904–916.

[40] CAIDA.(2016). *The CAIDA IPv6 Routed /48 Topology Dataset*. Accessed: Nov. 11, 2024. [Online]. Available: https://www.caida.org/catalog/datasets/ipv6_routed_48_topology_dataset

[41] F. Holzbauer, M. Maier, and J. Ullrich, "Destination reachable: What ICMPv6 error messages reveal about their sources," in *Proc. ACM Internet Meas. Conf.*, Nov. 2024, pp. 280–294.

[42] W. A. Simpson, D. T. Narten, E. Nordmark, and H. Soliman, *Neighbor Discovery for IP Version 6 (IPv6)*, document RFC 4861, Sep. 2007. [Online]. Available: https://www.rfc-editor.org/info/rfc4861

[43] R. Beverly, R. Durairajan, D. Plonka, and J. P. Rohrer, "In the IP of the beholder: Strategies for active IPv6 topology discovery," in *Proc. Internet Meas. Conf.*, Oct. 2018, pp. 308–321.

[44] E. Kohler, J. Li, V. Paxson, and S. Shenker, "Observed structure of addresses in IP traffic," in *Proc. ACM SIGCOMM Workshop Internet Meas.*, 2002, pp. 253–266.

[45] J. Korhonen, J. Arkko, T. Savolainen, and S. Krishnan, *IPv6 for Third Generation Partnership Project (3GPP) Cellular Hosts*, document RFC 7066, Nov. 2013.

[46] R. Padmanabhan, J. P. Rula, P. Richter, S. D. Strowes, and A. Dainotti, "DynamIPs: Analyzing address assignment practices in IPv4 and IPv6," in *Proc. 16th Int. Conf. Emerg. Netw. EXperiments Technol.*, Nov. 2020, pp. 55–70.

[47] T. Mrugalski et al., *Dynamic Host Configuration Protocol for IPv6 (DHCPv6)*, document RFC 8415, Nov. 2018.

[48] T. Narten, R. Draves, and S. Krishnan, *Privacy Extensions for Stateless Address Autoconfiguration in IPv6*, document RFC 4941, Sep. 2007.
[49] K. Fukuda and J. Heidemann, "Who knocks at the IPv6 door: Detecting IPv6 scanning," in *Proc. Internet Meas. Conf.*, Oct. 2018, pp. 231–237.
[50] J. Woodyatt, *Recommended Simple Security Capabilities in Customer Premises Equipment (CPE) for Providing Residential IPv6 Internet Service*, document RFC 6092, Jan. 2011.
[51] M. Ziv, L. Izhikevich, K. Ruth, K. Izhikevich, and Z. Durumeric, "ASdb: A system for classifying owners of autonomous systems," in *Proc. 21st ACM Internet Meas. Conf.*, Nov. 2021, pp. 703–719.
[52] C. Partridge and M. Allman, "Ethical considerations in network measurement papers," *Commun. ACM*, vol. 59, no. 10, pp. 58–64, Sep. 2016.

Tao Yang received the bachelor's degree in network engineering and the master's and Ph.D. degrees in computer science and technology from the National University of Defense Technology, China, in 2019, 2021, and 2025, respectively. His research interests include network measurement and network security.

Bingnan Hou received the bachelor's and master's degrees in network engineering from Nanjing University of Science and Technology, China, in 2010 and 2015, respectively, and the Ph.D. degree in computer science and technology from the National University of Defense Technology, China, in 2022. His research interests include network measurement and network security.

Yifan Yang received the bachelor's degree from the School of Cyberspace Security, Northwestern Polytechnical University, China, in 2023. He is currently pursuing the Ph.D. degree with the College of Computer Science and Technology, National University of Defense Technology, China. His main research interests include network security, network measurement, and network topology discovery.

Zhenzhong Yang received the bachelor's degree in network engineering from the National University of Defense Technology, China, in 2024, where he is currently pursuing the master's degree with the College of Computer Science and Technology. His main research interests include network security and network measurement.

Zhiping Cai received the bachelor's, master's, and Ph.D. degrees in computer science and technology from the National University of Defense Technology, China, in 1996, 2002, and 2005, respectively. He is currently a Full Professor with the College of Computer Science and Technology, National University of Defense Technology. His main research interests include network security and edge computing.