Fiat Lux: Illuminating IPv6 Apportionment With Different Datasets

AMANDA HSU, FRANK LI, and PAUL PEARCE, Georgia Institute of Technology, USA

IPv6 adoption continues to grow, making up more than 40% of client traffic to Google globally. While the ubiquity of the IPv4 address space makes it comparably easier to understand, the vast and less studied IPv6 address space motivates a variety of works detailing methodology to collect and analyze IPv6 properties, many of which use knowledge from specific data sources as a lens for answering research questions. Despite such work, questions remain on basic properties such as the appropriate prefix size for different research tasks.

Our work fills this knowledge gap by presenting an analysis of the apportionment of the IPv6 address space from the ground-up, using data and knowledge from numerous data sources simultaneously, aimed at identifying *how* to leverage IPv6 address information for a variety of research tasks. Utilizing WHOIS data from RIRs, routing data, and hitlists, we highlight fundamental differences in apportionment sizes and structural properties depending on data source and examination method. We focus on the different perspectives each dataset offers and the disjoint, heterogeneous nature of these datasets when taken together. We additionally leverage a graph-based analysis method for these datasets that allows us to draw conclusions regarding when and how to intersect the datasets and their utility. The differences in each dataset's perspective is not due to dataset *problems* but rather stems from a variety of differing structural and deployment behaviors across RIRs and IPv6 providers alike. In light of these inconsistencies, we discuss network address partitioning, best practices, and considerations for future IPv6 measurement and analysis projects.

CCS Concepts: • **Networks** → **Network measurement**.

Additional Key Words and Phrases: IPv6; Network Measurement; IP Allocation; IP Apportionment

ACM Reference Format:
Amanda Hsu, Frank Li, and Paul Pearce. 2023. Fiat Lux: Illuminating IPv6 Apportionment With Different Datasets. *Proc. ACM Meas. Anal. Comput. Syst.* 7, 1, Article 21 (March 2023), 24 pages. https://doi.org/10.1145/3579334

21

1 INTRODUCTION

Broad understanding of the division, allocation, and utilization of IP addresses is of the utmost importance for applications ranging from making security and policy decisions (e.g., blocklisting and equitable allocation), to measurement and analysis (e.g., address assignment practices), to generative IP tasks (e.g., efficient Internet scanning [26]). Such understanding of IP apportionment [1] is also critical to understanding if the established policies and procedures from governing bodies are reflected in the instantiation of those policies.

Many facets of the IPv4 space are well understood, largely driven by our ability to exhaustively probe and scan each address within the space [17]. Operators and researchers have a consistent understanding of basic IPv4 characteristics and have built widely used datasets around IPv4 measurement [10, 16, 23]. Our understanding of the IPv6 address space, however, is relatively nascent.

[1]We consciously choose a new term, "apportionment", to convey the high-level notion of division and assignment while avoiding conflation with or overloading existing terminology such as RIR allocation.

Authors' address: Amanda Hsu, ahsu67@gatech.edu; Frank Li, frankli@gatech.edu; Paul Pearce, pearce@gatech.edu, Georgia Institute of Technology, USA.

2476-1249/2023/3-ART21
https://doi.org/10.1145/3579334

Despite significant policy efforts, answers to even the most basic questions remain unclear, such as the right granularity to identify end-hosts [31]. Meanwhile, IPv6 adoption continues to grow, with a 20% increase in observed IPv6 end-host traffic in the last two years, totaling more than 40% of total traffic seen by online services [22].

The intersection of the growth of IPv6 with a lack of understanding of the apportionment of the IPv6 space represents significant challenges both in research and practice. Various studies use different heuristics (e.g., prefix length) and different datasets (e.g., RIR bulk WHOIS records vs delegation files) for exploring IPv6, resulting in potentially fundamentally different results [11, 19, 20, 33, 37]. This problem extends to practitioners as well, such as with debate over how to effectively identify and block IP-driven abuse [31].

Our work addresses this knowledge gap, providing insights on IPv6 space apportionment, IPv6 datasets, and considerations for future work in this space. We seek to understand how different publicly available datasets provide differing perspectives on the IPv6 address space and focus on how the address space is apportioned *in practice*, with an eye toward how those differences can impact measurement and analysis.

To achieve such illumination we focus on understanding apportionment from the ground-up, generating understanding by detailed examination of bulk WHOIS records [2, 6, 9, 30] and delegations files [1, 5, 8, 29, 42] from Regional Internet Registries (RIRs), advertised BGP routes from Route Views [35] and RIPE RIS [45], and active IPv6 addresses from hitlists [20] and other public sources we collect. While these datasets are fundamentally different, they have been used in various contexts for dividing up the IPv6 space. Intersecting each of these datasets is a forest-based analysis technique we introduce that leverages the allocation/suballocation relationship across RIRs to provide a deeper understanding of apportionment. We find significant variation in our understanding of apportionment based on the dataset explored and analysis technique, with each painting a different picture of the structure and utilization of IPv6 effective for different research goals. Based on this understanding we propose specific guidance for future IPv6-related tasks.

More specifically, our contributions include:

- Detailed review of how previous IPv6 work has leveraged WHOIS records and routes, including assumptions made about the datasets, address grouping, and size of active prefixes (Section 3).
- Detailed overview and characterization of all publicly available IPv6 data sources (Section 4).
- Examination of three distinct aggregate perspectives originating from RIR data: delegation files, bulk records, and RIR policies. We find none gives a full picture of IPv6 apportionment and all three are necessary to describe the dynamic deployment from the RIRs (Section 5).
- Based on our exploration of RIR data, we use a forest-based analysis approach for IPv6 apportionment that leverages the inherent structure of address space apportionment. We identify there is significant *heterogeneity* in structure according to the different stakeholders in the area of space, and that different areas of IPv6 space are fundamentally differently apportioned (Section 6).
- Examination of IPv6 apportionment from the perspective of BGP routing data. This includes the application of our apportionment-based method to routing data and how this enables us to use the data to find the largest *effective* prefixes of IPv6 space by observed route (Section 7).
- Exploration of the merging of RIR and BGP routing data together into a cohesive view of IPv6 apportionment. We leverage our forest-based method to determine characteristics that allow identification of groups of prefixes that are either entirely active or inactive (Section 8).
- Enhancing our comprehensive view of apportionment with active IPv6 addresses collected from numerous sources to identify groups of active prefixes based on forest-characteristics. We find that active addresses, routes, and WHOIS records intersect in highly variable ways across the space, demonstrating significant heterogeneity of usage and density (Section 9).
- Recommendations for future work and studies (Sections 10).

2 RELATED WORK

IPv6 has been studied across numerous dimensions. Here we outline the state of IPv6 research that shapes our understanding and considerations. We also consider how previous work has used the same datasets we do. We differentiate our work by emphasizing that we are studying IPv6 apportionment from a high-level lens in order to characterize how the space is used from the ground-up; prior work has measured adoption and utilization over time, IPv6 end-device address assignment, structural analysis within addresses, and aggregating active addresses.

IPv6 Adoption. Many works have studied how IPv6 has been deployed and adopted over time. In 2009, Karpilovsky et al. studied IPv6 deployment using delegation files and BGP routing data [27]. They found that while IPv6 was being deployed at an exponential rate, almost half of allocations did not have observable routing advertisements. However, they do not analyze any sub-allocation structure, and instead focus more on passive Internet traffic analysis to characterize IPv6 deployment. In our work, we instead focus on how to apportion the IPv6 space and include bulk WHOIS records, a fundamentally different dataset than the delegation files (Section 4).

In 2014, Czyz et al. [15] approached measuring IPv6 adoption using RIR data (using, we believe, delegation files), routing data, DNS data, and traffic analysis. They similarly find that IPv6 is being allocated at an exponential rate as well as characterize the types of applications using IPv6, and comment on IPv6 topology. However, in this work, the authors again focus on measuring adoption, whereas we focus on ways to break up the space (apportionment). However, the authors do find that adoption is not uniform over geographic areas, an idea that we explore deeper in Section 5.

IPv6 Assignment Practices. There are numerous works examining the IPv6 addresses that are assigned to end-devices. While we do not study end-device assignment, these works illustrate the dynamic and complex nature of IPv6 usage. Padmanabhan et al. temporally analyzed addressing assignment practices in both IPv4 and IPv6 [36]. They found that, motivated by IPv4 exhaustion, IPv4 address leases are significantly shorter than those of IPv6. They additionally find that this differs between Internet registries and connectivity providers. We consider their findings on the differences between IPv6 in RIRs as we use WHOIS records to characterize IPv6 deployment.

Rye et al. studied IPv6 address assignment practices and presented methodology that allowed tracking of individual IPv6 clients [48]. This work characterizes *prefix rotation* of Customer Premises Equiptment (CPE) within a larger prefix that belongs to a provider. We study apportionment from a more general lens and do not study assignment practices specifically.

IPv6 Address Structural Analysis. Prior work has leveraged the relationships between different parts of addresses (structural analysis) in order to draw conclusions about usage across the space. Plonka et al. developed methods of temporal and spatial classification of IPv6 addresses [37]. Their temporal analysis measures the stability of a prefix and subsequent addresses and their spatial analysis analyzes the structure of the prefix and its' addresses. They find several dense prefixes that contain high numbers of active addresses relative to their size. Additionally, they discover patterns in addressing for different datasets including ISPs in different geographic areas and mobile carriers.

Entropy/IP was the first machine-learning-based address generation approach [19]. It generated clusters of address nybbles (hexadecimal boundaries in IPv6 addresses) and used a Bayesian network to define the address probability distribution. It then generated addresses that were likely to be active. Similarly, Murdock et al. used clustering to group addresses and generate based on clusters [33]. Ullrich et al. also leveraged pattern-based analysis of IPv6 addresses in order to generate new addresses that are likely to be active [51]. We highlight again that we do not study low-bit patterns in our work and rather focus on the apportionment of IPv6 prefixes. We do, however, consider what these works have shown in terms of the higher bits of IPv6 addresses.

IPv6 Address Collection and Analysis. Other works have aggregated active IPv6 addresses from a variety of sources and methods to study deployment patterns across a variety of metrics. Beverly et al. explore ICMPv6 probing methodology that alleviates the effects of rate limiting on active measurements for IPv6 topology discovery and allowed them to discover active IPv6 addresses on router interfaces [11]. In 2018, Gasser et al. created the largest publicly available IPv6 hitlist [20]. This work was subsequent to earlier work by Gasser et al. that leveraged a variety of passive and active collection methods for IPv6 addresses in 2016 [21]. Subsequent follow-up work expanded the method and public hitlist [54]. A core contribution of this work is that the distribution of the hitlist had better diversity than previous work on IPv6 address collection. This was measured by comparing the addresses to advertised BGP routes and ASes that appeared in the Farsight DNSDB [18]. We use inspiration from these metrics to evaluate the distribution of our collected IPs. We use their publicly available hitlist as a part of our dataset of active addresses.

WHOIS Records and Routes Analysis. Previous work has also compared data from RIRs to routing data. Nemmi et al. analyzed bulk WHOIS records alongside advertised routes [34]. However, their study focused on discrepancies between autonomous system number (ASN) allocations and observed routes over 17 years, not IP ranges. Their findings identify many temporal errors in records and routes such as routes being advertised before their ASN allocation record appears, or ASNs in two records at once while its ownership is being transferred. We consider these errors relevant in our analysis of the intersection of IP ranges and advertised routes. However, we focus on the apportionment of the IP ranges in records and do not study ASNs in RIR data.

3 USE OF IPV6 APPORTIONMENT AND DATASETS IN PRIOR WORK

We now elaborate on prior work utilizing the datasets we examine, with an eye towards understanding how the choice of dataset may impact the work and its understanding of apportionment. We group work by their apportionment *goal* rather than their dataset. We reiterate that assumptions about how to use datasets such as WHOIS records and routing data motivate our work.

Using Datasets as Grouping Heuristics. A number of works have used BGP datasets, RIR policies, and assumptions about IPv6 address space partitioning. These studies group addresses together at certain granularities, with the underlying assumption being that addresses within a certain unit (e.g., route, prefix) will have the same address assignment policy and patterns. For example, 6Gen [33] used Route Views to group seed addresses together in routed prefixes, explicitly assuming that patterns in active addresses would manifest at that granularity.

Similarly, in their hitlist analysis, Gasser et al. used BGP data to group addresses in order to perform entropy analysis to identify similar addressing patterns within a routed prefix [20]. They also evaluate the diversity of a hitlist across autonomous systems at the routing granularity, comparing their hitlist to prior works along a routing-based metric.

Beverly et al. also used BGP data to generate a control seed list by selecting random targets within advertised prefixes [11]. They used different prefix lengths to group addresses, accounting for different prefix lengths that different providers may use. In their generation methodology, they apply a prefix transformation to their input prefixes in order to maintain a constant prefix length. The authors comment on the importance of choosing the granularity at which they do this and note inconsistencies with expected active addresses within /64 boundaries. Specifically, they find that multiple traceroutes to different addresses within the same /64 prefix find different topologies. In this work, we explore different structural dimensions of such datasets, evaluating and hypothesizing about whether these assumptions about grouping are effective in IPv6 research goals.

IPv6 Prefix Size Assumptions. Other works use certain static prefix lengths for grouping, particularly the /32 prefix length due to RIR policy in assigning subnets of this size to providers.

Plonka et al. use this assumption in their analysis of active IPv6 addresses to group addresses in prefix lengths between /32 and /64 (as providers with /32s may sub-apportion); however they acknowledge that many RIRs allocate larger prefixes than /32 [37]. Similarly, in their work on using IPv6 address structure for generation, Foremski et al. assume that the largest allocation to a provider is a /32 subnet. As a result, they treat the first 32 address bits differently during their entropy analysis [19]. However, as acknowledged by Plonka et al., many allocations to providers exist at large sizes. Murdock et al. also applied prefix heuristics when studying aliased regions of the IPv6 address space, assuming that aliased netblocks would not be assigned subnets smaller than a /96, and applying their detection technique at that prefix size. In this work, we explore assumptions about prefix lengths given to providers and make recommendations for future work to use a more flexible prefix length. We also hypothesize this will be increasingly important as IPv6 deployment grows, an idea that is reinforced by RIR policies that we discuss in Section 5.3.

4 DATA SOURCES

We now describe, characterize, and identify challenges in the four core datasets underpinning this (and prior) work: bulk WHOIS records from Regional Internet Registries (RIRs), delegation files from RIRs, routing data, and active IPv6 address sources. Table 1 provides a dataset overview.

RIR (Totals)	ARIN	RIPE	AFRINIC	APNIC	LACNIC
Delegations	69,157	104,165	5,936	92,965	50,940
Bulk Records	278,847	783,015	31,690	77,800	-
Delegated Space	$2^{112.0}$	$2^{113.4}$	$2^{109.3}$	$2^{112.6}$	$2^{110.0}$
Bulk Record Space	$2^{117.0}$	$2^{117.0}$	$2^{116.0}$	$2^{116.0}$	-
Routed Space	$2^{113.2}$	$2^{112.1}$	$2^{109.2}$	$2^{111.4}$	-
Routes	217,294	55,972	2,177	72,540	-

Table 1. Overview of the datasets used in this paper. The delegations and bulk records list the number of IPv6 delegations or records in each dataset, respectively. The delegated space, bulk record space, and routed space show the number of IPv6 addresses that are contained in each dataset. The routes list the number of IPv6 advertisements in the dataset. Note that delegated records include reserved and available ranges listed in the records, but only assigned and allocated space is included in the volume calculation.

4.1 Background on RIRs and Routes

Organizational Aspects. The Internet Assigned Numbers Authority (IANA) is the governing body that controls IP address space on the Internet [24]. IANA delegates blocks of IP addresses to Regional Internet Registries. Each of the 5 RIRs—ARIN (North America), APNIC (Asia), RIPE (Europe), LACNIC (Latin America), and AFRINIC (Africa)—represent a distinct geographic region and are responsible for further suballocation of IP addresses within their region [24] to other parties. For example, IANA may allocate a large range of IPs to ARIN, and, in turn, ARIN may suballocate some of those IPs to an ISP, who may in turn again suballocate those IPs to individual business customers. Each RIR has its own policies for how their respective IPv6 space can be apportioned, and as such maintains records in different ways. Although the policies are largely similar, the specificities of records differ slightly. RIRs are simply one type of Internet Registry (IR); National Internet Registries (NIRs) may manage more specific geographic regions within the RIR. The policies of the RIR may or may not apply and be carried out by the NIR. There also exist Local Internet Registries (LIRs) that apportions address space most directly to the user [4].

WHOIS Records. Each of the 5 RIRs maintains detailed WHOIS records documenting the management of all IP address in their purview. WHOIS records are intended to serve as the primary record of "ownership" of each IP range on the Internet, with each record documenting the mapping of a continuous IP range to the owner. WHOIS records are the publicly available resource that most closely reflects the ground truth of IP address status, containing direct allocation and assignment of IP addresses from RIRs as well as any suballocation and assignment by those entities. However, previous work has found that they may contain temporal errors, such as when ownership of a resource is being transferred [34].

Record Types. When an RIR delegates an IP range, it can (generally) have one of three possible states: allocated, assigned, or reserved (within RIRs, there can be other terms such as reassignment, policy-reserved, and other variations). Assigned IPs are delegations to an organization that cannot be sub-allocated (broken down) further. Allocated IPs permit the receiving organization to subdivide the range further. Reserved IPs are ranges that have a special purpose as decided by either the RIR or IANA. A *direct* allocation or assignment is when the IP range is given to the organization directly from the RIR, not another organization (like a NIR or LIR).

Routes. The IPv6 BGP routing system functions similarly to IPv4, and prior work has studied the specificities of these topologies [47]. RIRs recommend that IPv6 routes are recommended to be of a minimum length /48, but previous work has measured advertisements with a prefix length of 49-64 propagating through the global routing system [50, 53] as well as more specific prefixes observed by collectors [49]. In total, we filtered out 164 routed prefixes that were more specific than /64.

4.2 Data From RIRs

We examine two IPv6 apportionment datasets from Regional Internet Registries (RIRs): bulk WHOIS data and delegation files. These datasets both describe aspects of IPv6 allocation but are fundamentally different in function and structure. We use snapshots from August 1st, 2022.

Bulk WHOIS. In order to facilitate research, each RIR provides research access to their WHOIS records in bulk. Bulk records additionally include more detailed information on each organization that owns any record. This includes contact information (e.g., email and phone number). Many organizations also include comments on their assets and what to do in case of emergency or abuse. *Our Dataset:* We obtained bulk WHOIS information from each RIR as the starting point, which we then intersect with other data sources. Our dataset consists of 1,171,352 records across 4 of the 5 RIRs from August 1st, 2022. Unfortunately, we were unable to obtain the bulk WHOIS information from LACNIC (Latin America); they are excluded from our study. [2] After obtaining bulk records some sanitization was necessary. Specifically, we omit records pertaining to reservations (a special type of record for a variety of RIR purposes, one of which is to temporarily "hold" space near an existing organization in the event that an organization grows) and remove any duplicate IP ranges across RIRs to prevent double counting.

Delegation Files. Each RIR also has a public dataset of delegation files [1, 5, 8, 29, 42]. Delegation files, to a first approximation, contain less information about sub-apportionment of IPv6 space. While the bulk records may contain information about many levels of suballocation or assignment, delegation files may only contain direct allocations and assignments from the RIR or only one additional level of sub-apportionment. These files include information on the ASs and ranges of IP addresses that are managed by the RIR and are published daily. The data is in a standardized format for all RIRs. For each IP range, the file specifies the registry it belongs to, the country code of the organization that owns it, the type (IPv4 or IPv6), the IPv4 address range or an IPv6 prefix, the date of the record, the status of the delegation (one of: allocated, assigned, reserved, available), and, optionally, extra data the RIR can specify. Additionally, there is an identifier called an opaque ID for each record that associates it with an organization. The organization details are not disclosed, but the IDs can be used to group delegations together that belong to the same organizations. *Our Dataset:* Using the publicly available FTP servers, we obtain delegation files from the 5 RIRs consisting of 323,163 records, again on August 1st, 2022.

Differences Between Delegation Files and Bulk Records. While both RIR datasets provide a view of address apportionment, they vary in both their granularity (i.e., level) and statuses. Bulk RIR

[2]We attempted to obtain LACNIC's dataset numerous times. After several engagements, we could not obtain the data.

WHOIS records represent a tree of IP address apportionment. e.g., if an RIR assigns an IP range to Org. A who in turns suballocates part of their range to Org. B, and so-on, the structure is captured in WHOIS records with IP ranges and organizational information. In contrast, delegation files typically only capture the exchange between the RIR and Org. A—the first level of delegation. Section 5 shows the behavior of IP address apportionment varies significantly with level, necessitating the need for a more fine-grained view of RIR behavior than afforded by delegation files.

In Table 1 the space described in delegations is orders of magnitude less than that of the bulk records for each RIR. This is due to allocations listed in the delegations, not including the RIR allocations (from IANA or another RIR). In our calculations of the delegations space we only include allocated and assigned space, not available and reserved. However, including the available and reserved space in this calculation results in the same delegation space as the bulk record space.

4.3 Routing Data

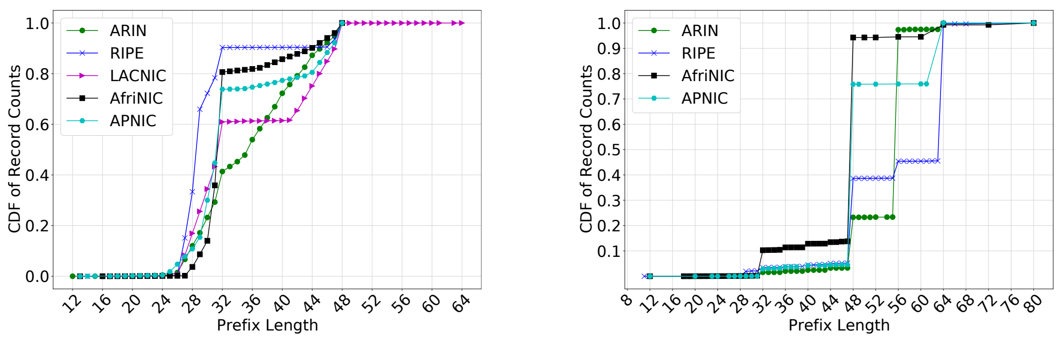
We collect advertised BGP routes from Route Views [35] and RIPE RIS [45] datasets. RIPE NCC and Route Views are public sources of routing data [35, 45]. They collect both BGP updates and full Routing Information Base (RIB) with various collectors that peer with volunteers in a variety of geographic regions. We use all the RIB files collected on August 1st, 2022, from each IPv6 collector in both data sources. Table 1 presents an overview of the data. We parse out and use the prefixes advertised from the RIB files. For sanitization, we remove routes less specific than a /8 and more specific than a /64, following the best practices of prior work [34]. We further exclude IANA reserved ranges, and special use ranges.

4.4 Active IPv6 Addresses

We obtain the aggregated IPv6 addresses from Gasser et al. [20, 54] as a starting point for active IPv6 addresses. We further collect domain names from Cisco's Umbrella [52], CAIDA's DNS [12], Rapid7 Forward DNS [38], and Majestic One Million [32] datasets. We then add to these the domain names collected from all X.509 certificates seen by Censys [16]. We then perform DNS IPv6 resolutions of all of these domain names from university network on the west coast of the United States and add all resulting IPv6 addresses to our active address dataset. We also include addresses from CAIDA's Ark IPv6 Topology Dataset [13], RIPE Atlas [43], and the (deanomymized) WIDE [14] dataset. In total, we collected 35,007,374 IPv6 addresses. Our goal with the collection of these (potentially formerly) active addresses is not to produce a new comprehensive hitlist but rather to create as vast as possible set of addresses to overlay against our IPv6 WHOIS and routing datasets.

5 IPV6 APPORTIONING FROM RIR DATA: THREE AGGREGATE VIEWS

We begin by evaluating the apportionment of the IPv6 address space from RIR WHOIS data, which we can consider the ground-truth assignment, allocations, and suballocations of address space regions to different networks and organizations. As mentioned in Section 4.1, previous work [34] identified many temporal errors in RIR data, however, evaluation of potential errors in this dataset is out of the scope of this work. We, therefore, characterize the apportionments of the entirety of possible IPv6 space. We ask the most basic IPv6 partitioning question: "What prefix lengths are most common in records?" To answer this question we explore three distinct perspectives on the aggregate prefix sizes in WHOIS data: 1) delegation files (Section 5.1), 2) bulk records (Section 5.2), and 3) RIR policies (Section 5.3). Sections 5.1 and 5.2 describe what the WHOIS datasets convey about ground-truth apportionment, whereas Section 5.3 describes what policies tell us IPv6 apportionment should theoretically be in broad terms. We include these three sections in order to compare the three different potential expectations for apportionment.



(a) Delegation Files.

(b) Bulk Records.

Fig. 1. CDFs of Record Prefix Lengths across RIR Datasets. (a) shows the most common *delegation* prefix length is /32, and prefixes of smaller lengths (and larger volumes of addresses) make up between 29.6%-78.1% of records (2,389 to 79,421 total records). (b) shows the most common *bulk* prefix length is /48, and prefixes of smaller lengths (and larger volumes of addresses) make up between 3.2%-13.5% of records (1,338 to 38,332 total records). x-axes are truncated.

We find that delegation files give insight into the direct records from the RIR to their direct customers (e.g., ISPs and others who receive direct records from the RIR). By analyzing the most common sizes of these records, we conclude that these are mostly large service providers. We find that the bulk records offer more insight into a finer level of apportionment, showing suballocations and assignments from these large providers. However, their aggregate statistics are more likely to be subject to bias from large providers. RIR policies provide commentary both on the expectations for the aggregate statistics relating to both providers and smaller suballocations and assignments. These policies also begin to explain the outliers in both of these datasets that we observe. From this evaluation, we conclude **each contributes a unique and incomplete perspective on the dynamic nature of IPv6 apportionment.** To better capture the incomplete interplay between these datasets, we develop a forest-based analysis technique (Section 6).

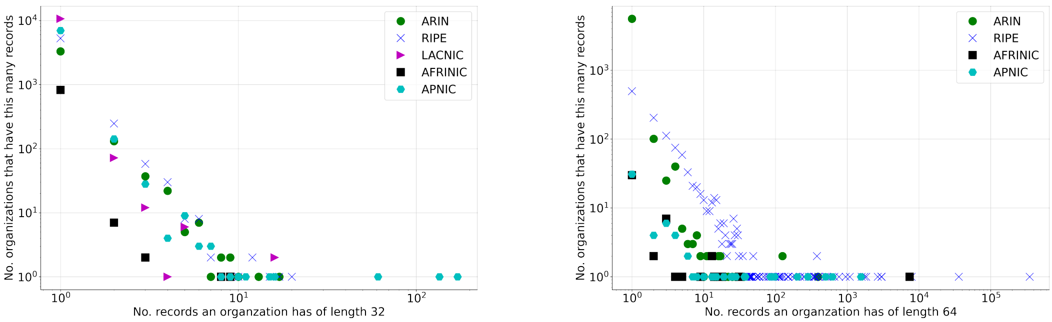
5.1 Delegation Files

Across all delegation file records, we find that the most common prefix length is a /32 (19.1% of records), followed by /29 (15.0%), /28 (10.1%), and /31 (9.5%). This distribution varies greatly by RIR (Figure 1a). In RIPE's data, the most common prefix is a /29 (32.6%). For the other RIRs, a /32 prefix is most frequent, but the proportion of records that are /32 prefixes differs. For ARIN, only 12.1% of its records are /32 prefixes, whereas it is 17.6% for APNIC, 29.1% for AfriNIC, and 44.7% for LACNIC. From these numbers, the apportionment of IPv6 space would *appear* primarily guided by large providers being delegated /28-/32s. From this one could (incorrectly) therefore conclude that IPv6 apportionment is largely driven by providers and their allocations directly from the RIR.

This view, however, is significantly incomplete; we find (discussed next) that, from the perspective of the bulk WHOIS records, **the driving form of apportionment is suballocation and assignment of the larger records of providers to other entities**. We note however the utility of each perspective may vary by the use case of the data, explored further in Section 10.

5.2 Bulk Records

Across all RIR bulk records, the most common prefix length is a /64 (37.7% of records), followed by a /48 (34.4%) and a /56 (22.1%). This distribution is in stark contrast to the delegation files, where



(a) Delegation Files. (b) Bulk Records.

Fig. 2. We show bias from individual organizations in bulk records and that there is little to no bias from individual organizations from delegation files. (a) shows that for records of prefix length 32, there is not much skew in any RIR besides APNIC, which has three organizations with over 50 records of this length. (b) shows there is a bias from individual organizations, especially in RIPE.

large subnets such as /32s were most frequent. This difference arises as the bulk records include more granular (and thus typically smaller) address space assignments and allocations, including those intended as assignments to individual end-sites. As shown in Figure 1b, we also observe variation in the most frequent prefix length appearing in the bulk records *between each RIR*. For ARIN, the most common prefix is a /56 (74.0%), whereas, for RIPE, it is a /64 (54.1%). The /48 is most common for APNIC (80.5%) and AfriNIC (71.2%). [3] With knowledge from the delegation files, this analysis implies that the IPv6 space is significantly suballocated and assigned in certain cases. We can additionally conclude that, **as opposed to the delegation files, the bulk records offer insight to suballocation and assignment within providers.**

We also note that the assignments intended for end-sites vary between /48s, /56s, and /64s, and thus the differences between RIRs may reflect varying policies on end-site assignments. Best practice policies for IPv6 prefix assignments state that prefixes longer than /56 are strongly discouraged [46], yet such assignments still exist. However, like many RIR policies, there are reasonable explanations for using more specific prefixes for assignments as long as they are not more specific than a /64. We discuss RIR policies in depth in Section 5.3.

A caveat of this form of aggregate analysis is that it can be biased by individual organizations and customers of the RIRs. Figure 2 shows that this occurs in bulk records. We highlight that the highest numbers of /64 records owned by the same organization within RIPE are 36,010 and 349,612. Figure 3 explores this challenge, showing that there are a significant number of unexpected RIPE records of length /64 starting in 2016. We observe that such trends in specific organizations can overwhelm the data at the aggregate level, motivating a more advanced analysis technique (Section 6).

5.3 RIR Policies

RIR Policies should give us general expectations on our aggregate statistics for both delegation files and bulk records. The flexible nature of these policies should also explain any variation observed.

In all cases, the maximum allocation size for IPv6 space is a /32, but policy allows for exceptions in which larger allocations can be made [7, 28]. Figure 1a shows such exceptions; most records

[3] As discussed in Section 4.2, we could not obtain LACNIC bulk records. We omit further reiteration of this challenge.

in the delegation files are size /32, but also *significant* numbers of prefixes are larger and smaller lengths, highlighting the diversity in the RIR policies and analysis challenges.

End-site assignment is at minimum a /64 and at maximum a /48, although a /48 is recommended in most cases; this decision is ultimately left to the LIRs. Flexibility is encouraged to promote best practices relative to the structure, services, and other characteristics of individual providers [4, 28, 44]. Figure 1b shows this, as we observe each RIR has a majority of records of one of these lengths, but again with variation due to specific customer decisions.

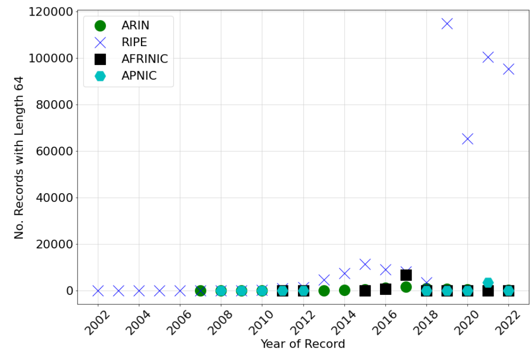


Fig. 3. Aggregate statistics of bulk WHOIS records of size /64. We show the number of records of size /64 that are created each year. We find that a handful of organizations allocating at this size can overwhelm aggregate analysis.

We can also use RIR policies to reason about the trends we observe in delegation files and bulk records (Sections 5.1 and 5.2). Although RIRs tend to initially allocate in /32s, a customer's apportionment can be increased in size if they are utilizing the space effectively and need more room to grow their network. One metric that RIRs use to evaluate this is the HD Ratio:

$$HD_r = \frac{\log(\text{number of objects allocated/assigned})}{\log(\text{number of objects that could be allocated/assigned})} \tag{1}$$

This ratio was originally described in RFC 1715 and subsequently updated and discussed in RFCs 3194 and 4692 [39–41]. The HD Ratio exists for the purpose of measuring utilization in a non-linear fashion for IP addresses, phone numbers, and other address-driven number management.

The numerator is typically calculated with respect to a /48 or /56 allocation depending on the customer type. For this reason, RIRs require that their customers register their assignments. RIRs maintain a flexible approach to this evaluation and also allow their customers to provide other documentation with rationale for an increase in space. We highlight that, by using assignment records in this calculation, RIRs are not actively or passively measuring utilization. Rather, they are measuring intention by various organizations. For IPv4, the HD ratio that an RIR requires for a larger allocation is 0.8. The same was required for IPv6 until 2007 when AFRINIC implemented a change in policy to increase the HD ratio from 0.8 to 0.94 for IPv6. [3].

RIPE's policies state that instead of using the HD Ratio, an LIR can describe its needs and usage in terms of its number of users, specificities about its infrastructure geographically and hierarchically, for security purposes, or prove that it will preserve the longevity of the original allocation [44]. For this reason, any /48 that is assigned to an end-site must be registered with RIPE. During this process, if a prefix that is less-specific than a /48 is used for a single user or end-site, the LIR must justify it. This is true across RIRs [4, 7].

Figure 1 shows that increasingly organizations are moving towards large "unconventional prefixes," possibly due to an increase in adoption of IPv6 [22] and therefore an increase in their initial /32 allocations or simply strong justification that they need this size of allocation to start, although for the most part apportionments stay at conventional sizes.

5.4 Observations Across Perspectives

Recall from Section 4.2 that the delegation files are largely a subset of the bulk records. We can describe the perspective that the delegation files provide with respect to WHOIS structure. In ARIN's data, *all* IPv6 assignments and allocations (8,472 total) listed in the delegation files are on level 1 of ARIN's structure, which means it lists the apportionments directly from the RIR. In RIPE's delegation files, however, only 21,278 apportionments are on level 1 (85.7%), and 3,549 are on level 2 (14.3%), showing a higher granularity of apportionment. In AfriNIC's delegation files, every apportionment is also from level 1 (1,170). In APNIC's delegation files, 12,544 apportionments are from level 1 (99.9%), and 17 are from level 2 (0.001%).

From the differences in prefix length distributions across RIR datasets, we find that each provides different visibility into IPv6 apportionment. Further, even RIR policies themselves paint a different apportionment picture. The prevalence of longer prefixes in the bulk data suggests that many end-site assignments/allocations are not included in the delegation files, and instead delegation files primarily provide information about top-level allocations to large providers (directly from RIRs in many cases). Even in this case, we observe that large address allocations (in the delegation files) do not consistently align with the standard /32 prefix allocation that has been assumed from prior work (Section 3). **For understanding IPv6 apportionment at deeper levels of assignments and suballocations, we conclude that bulk WHOIS data is most appropriate.** However, we emphasize that the contribution from each data source is significant in painting the dynamic picture of IPv6 apportionment. These dynamics motivate our usage of a forest-based analysis technique able to more effectively capture apportionment dynamics, discussed subsequently.

6 FOREST ANALYSIS: CAPTURING STRUCTURE IN IPV6 APPORTIONMENT

Our analysis of the RIR WHOIS datasets focused thus far on the aggregate distribution of address assignments/allocations for each RIR. However, WHOIS records also afford visibility into the structure of the address space distribution that arises from multiple levels of space assignments and suballocations between various organizations. Based on this observation we use a forest-based method of analyzing the inherent structure of IPv6 WHOIS records. We use it to show the *heterogeneity in apportionment* by evaluating the characteristics of the forest. This allows us to characterize IPv6 apportionment when considering both the structure of assignments and suballocations. We apply this forest approach to the bulk WHOIS RIR dataset as the delegation files do not provide (sufficient, reliable) information about dependencies between records and ranges, and also lack visibility into finer-grained allocations/assignments.

6.1 Approach

RIR	Components	# Components w/ Children	Max Tree Depth	Max Tree Breadth
ARIN	9	6	6	208,255
RIPE	14	13	16	522,687
APNIC	9	5	5	60,529
AfriNIC	2	2	4	30,292

Table 2. Forest structure of the bulk WHOIS records for each RIR. We list the number of distinct components, the number of components with children (forming a tree), and the maximum depth and breadth observed for tree components. These statistics highlight structural differences in RIR apportionment.

WHOIS records form dependencies, as one record's IP range is an assignment or suballocation from that of another WHOIS record (unless a record is for a direct allocation from IANA or an allocation between RIRs). Such dependencies can be instantiated as a tree of records. An edge is defined from a *parent* record to a *child* record corresponding to an organization receiving an assignment/suballocation from a larger record.

By analyzing WHOIS records we observe that there are no cycles in such a graph (i.e., a child record assigning/suballocating addresses back to a parent record) and each child has one parent (i.e., no record's IPv6 range is assigned/suballocated from multiple other record's ranges). Thus, our structure is made up of direct acyclic trees. As there is no common root (discussed subsequently), such relations form a *forest* of directed acyclic trees, known as a forest.

To understand this structure in IPv6 address space distribution, we construct this forest for each RIR and evaluate its characteristics. We confirm that for all RIRs, the forest does not form a single tree, but rather it consists of multiple unconnected components, each of which is a tree. The root of each component's tree is an allocation to the RIR itself, which is either directly from IANA or potentially an allocation provided from another RIR, as mentioned. The rest of the component represents assignments or suballocations from within the root RIR's allocation. Note that in some cases, we observe root RIR allocations that have not been further assigned or suballocated, in which case the component is a singleton WHOIS record.

Initial Analysis. In Table 2, we list the number of distinct components and the characteristics of the component trees for each RIR. We choose the subsequently defined metrics as they are indications of apportionment structure. Comparing these high-level characteristics across the various RIRs allows for the identification of broad differences in their apportionment strategies. The number of components indicates, broadly, the initial allocations to the RIR. The number of components with children identifies how many of these initial allocations are being sub-apportioned to the RIR's customers. Meanwhile, the max tree depth reveals the maximum number of times an initial allocation has been sub-apportioned. Similarly, the max tree breadth shows the maximum number of times the same apportionment has been divided.

From this, we see varying structure across RIRs, with RIPE having the most at 14 components, whereas AfriNIC has only 2. RIPE also exhibits the most complex components, having tree components with more than twice the depth and breadth of any other RIR components. From this, we can infer that providers in RIPE's regions have more complex infrastructure than other regions with less depth. Larger providers may tend to suballocate their space to smaller providers, rather than smaller providers gaining IPv6 address space directly from the RIR.

By examining components in detail we observe that the tree structure is highly heterogeneous, demonstrating that there is not a uniform policy by which organizations or networks assign and suballocate their address space. Rather, each network provider applies different policies and practices in distributing their own address space. We identify this by analyzing the characteristics of our forest and observing that all sub-trees have significantly different depths and balances.

6.2 Structure Heterogeneity

We analyze one component of ARIN's IPv6 forest that is informative and representative to explore apportionment *heterogeneity*. **We observe significant variance due to providers having many different tendencies with their initial allocations and that the records intended for "end-site" usage occur at various levels within the tree.** We emphasize the importance of these considerations as they indicate areas of interest for future IPv6 studies.

Figure 4 depicts the distribution of prefix lengths at varying tree depths for one component in ARIN's WHOIS records. We focus on prefix length with respect to the tree depth because its heterogeneity shows that there is no generalizable rule for sub-apportionment. That is, the number of times an address space has been sub-apportioned does not dictate the size of the apportionment and therefore a variety of prefix sizes occur at every level of the tree. From the perspective of prefix lengths, we see a large range of sizes on each level of the tree. Each level represents the number of times the space from the original prefix has been suballocated or assigned out. We highlight that this is not a count of how many times the original prefix has been split up. For example, if 200 /48s have been allocated from a /32, we count the level of the /48 as 1, not 200.

At the second level of apportionment (direct ARIN allocations), we observe that the median apportionment size is /32. However, there is a large range of apportionment sizes, from a /20 to a /64. At deeper levels, we observe the median apportionment size does generally increase (e.g., at the third level, the median is a /48), but there still is high diversity (at lower levels, the diversity appears to decrease largely because the number of records at that level is limited). Thus, there is high variance in the assignment and suballocation behavior of different organizations, resulting in varying prefix sizes at each level of allocation regardless of number of times it is suballocated.

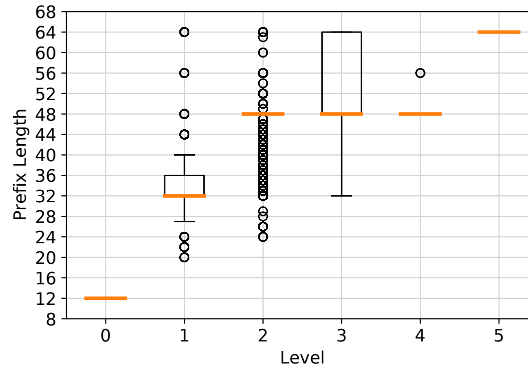


Fig. 4. Box and whisker plot of prefix length distributions for an ARIN component, across sub-apportionment levels. The colored line represents the median prefix length. We highlight the variety of ways at which space from previous levels is apportioned on the next level, indicating variation in provider and RIR behavior alike even within the broader address space.

Beyond examining the size of the prefixes, we now evaluate the relationships between prefixes, as depicted in Figure 5. In Figure 5a, we show the level of a given prefix size in a component to identify the relationship between a prefix size and the number of times the space has been sub-apportioned. In Figure 5b, we plot the size of a child record with respect to the size of its parent record, assessing the apportionment sizes that are sub-apportioned. In Figure 5c, we display the level of a record's sub-apportionment given its size, revealing the relationship between a prefix size and the number of times a space has been sub-apportioned. In Figure 5d, we investigate the relationship between a record size and the size of its parent, in order to understand how records are sub-apportioned in a heterogeneous fashion.

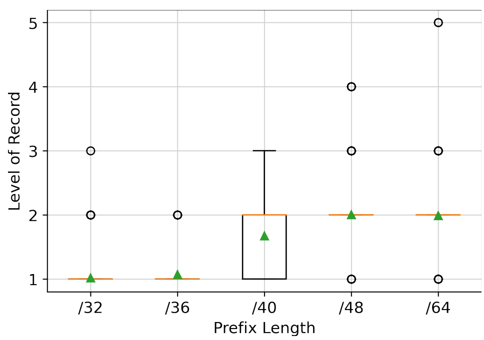
In Figure 5b we observe that, although there is low variance at the level of the children, there are large ranges. We highlight that this indicates that there are many outliers in the level at which children of a certain size exist. We observe in Figure 5c that lengths /32 and /36 have the most variance in the size of their children. We conclude that this is due to providers apportioning their space in a variety of ways.

We emphasize that the number of records on each of these levels varies and indicates diversity in provider sub-apportionment of their initial apportionment. The top record on level 0 is a /12 record with 5,548 children on the next level. 444 of the 5,548 records on level 1 have at least 1 child on the next level. On level 2, there are 56,984 records, 58 of which have children. On level 3, there are 667 records, 11 of which have children. On level 4, there are 27 records, 4 of which have children. Level 5 is the last level and has 4 records in total. We once again emphasize that providers apportionment in different ways at different levels of sub-apportionment.

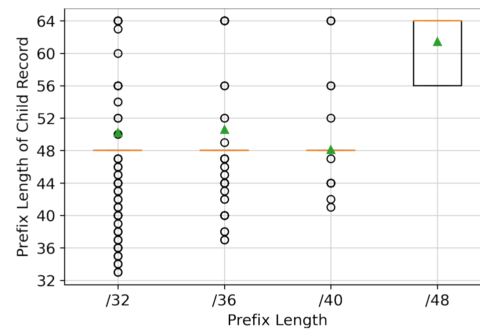
Our presented approach frames our understanding of IPv6 apportionment from RIRs and other data sources going forward. We use this as the basis to build our characterization of how apportionments from records differ from and intersect with apportionments from other data sources.

7 IPV6 APPORTIONING FROM ROUTING DATA

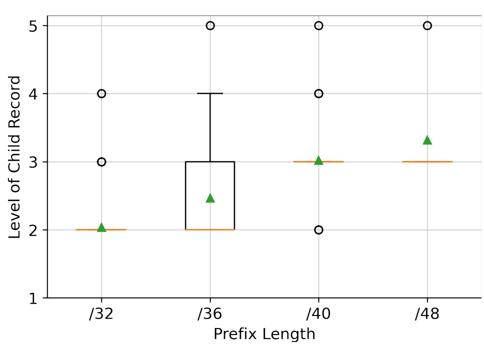
We next turn our attention to IPv6 address space distribution from the perspective of BGP routing data. While WHOIS data indicates what address regions have been allocated to an organization for potential use, routing data indicates the address regions that are actually reachable on the IPv6 Internet, thus providing a different perspective on address space apportionment.



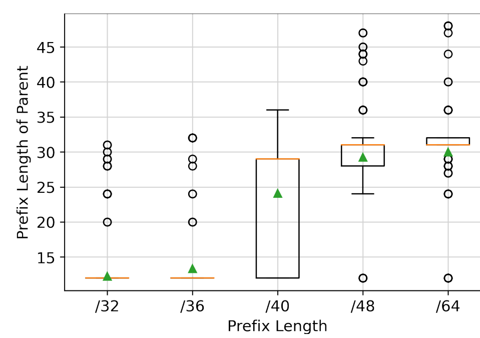
(a) Level of a given prefix size in the component.



(b) Size of a child record of given prefix.



(c) Level of a record's child records.



(d) Prefix length of the parent of a record.

Fig. 5. We analyze the top 5 most common prefix lengths in the ARIN component. We highlight that records of size /64 do not have children and are therefore omitted from 5b and 5c. 5a shows that although concentrated at certain levels, there is variance in where prefixes of a certain size appear in the tree. We highlight that records of size /64 appear at every level. 5b shows that the size of a prefix's child records varies. We emphasize the large variance in provider sizes (/32, /36, /40). 5c shows that, in addition to the size of a record's children, the level at which the record's children exist varies significantly. 5d shows the prefix lengths of the parent records, revealing the variance of this characteristic across the different record sizes.

Aggregate Perspective. We characterize route prefix sizes across all routes in Figure 6 (recall from Section 4.3 we filter routes for subnets smaller than /64). We observe that the most prevalent route prefix length is a /48 (46.3% of all routes), with /32s (13.5%), /44s (7.8%), and /40s (7.1%) also common. These common prefix sizes align with the recommended minimum route size of a /48 described in Section 4.1 (although nearly 10% of routes are smaller than RIR guidelines). The prevalence of /48 routes (as well as /44s and /40s) also indicates that the default initial apportionment of a /32 prefix to RIR customers is not typically routed in its entirety; in many cases, a sub-apportionment is routed, and only a fraction of a customer's space is *actually* potentially utilized.

Forest Analysis. While routing data is characteristically different from WHOIS data, there remains a notion of relationships between different routes, where one *child* route can be for a subnet (i.e., longer prefix) of a *parent* route. As we did with RIR WHOIS data (Section 6), we investigate the structure of route dependencies by constructing a directed forest, with edges between parent-to-child routes. With this forest approach, we identify that, unlike with WHOIS data, there are a significant number of unconnected components relative to the data size. There are 65,738 components for 643,156 total routes. This is reasonable as there are not a small set of overarching *source* routes from which most other routes are sub-prefixes. Instead, there are a large number of source routes, some of which have children routes. Table 3 shows the characteristics of the forest structure, demonstrating the diversity of the route components.

Components	Components w/ Children	Median Breadth	Median Depth	Max Breadth	Max Depth
72,876	63,800	2	2	4,289	8

Table 3. Routes parsed into forest representation. We note while in some components there is a maximum depth of 8 levels of sub-apportionment, most components only have 1 child, and are thus less complicated and more disjoint than that of RIR data.

To understand the relationship between route prefix lengths and whether they are source routes versus children routes, we plot the CDF of route prefix lengths in Figure 6 for source routes, children routes, and source routes with children. For both source routes, children routes, and sources without children, we observe that the most common route is a /48 subnet (46.3% of source routes, 44.2% of children routes, 52.7% of source routes with children). The prevalence of /48 prefixes among children routes suggests that their parent routes should skew towards larger subnets. We indeed observed this effect when considering the distribution of prefix lengths for only source routes without children, as also shown in Figure 6, where the most prevalent prefix is a /32 (59.8%). Thus these source routes, likely large network providers assigned /32 subnets who assign/suballocation their space to other organizations for separate routing, exhibit distinct routing behavior compared to other routes.

From this comparing different groups of routes with respect to our forest approach, we conclude that if providers route sub-apportionments of their space, it has different characteristics (i.e., sizes) than the overall routable space. This is driven by the structure within specific providers and is a result of a complex global routing system that is out of the scope of this work. However, we can characterize apportionments of routable space differently than the aggregate perspective on routing data. From this distinct perspective, we note the need to generate an aggregate view that leverages both RIR WHOIS data and BGP routing data. We explore such combined perspectives next.

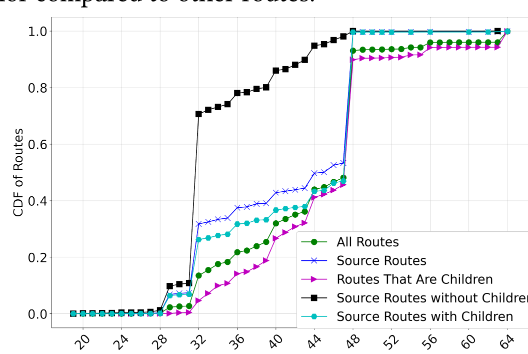
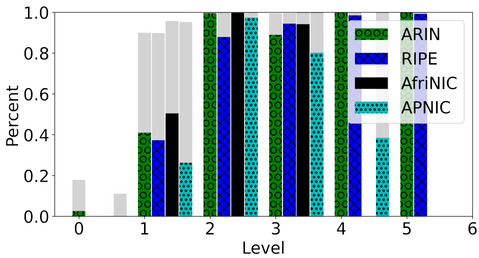


Fig. 6. Prefix Lengths for source routes and all routes. We observe that routes tend to be of size /32 or /48, largely depending on whether the route has children.

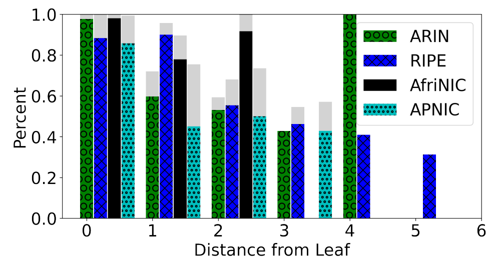
8 IPV6 APPORTIONING FROM COMBINING RIR AND ROUTING DATA

While records give us an indication of all possible IPv6 addresses, routing data indicates which subset of this data is potentially reachable. We thus improve our understanding of IPv6 apportionment by combining the RIR WHOIS records and routes, identifying the extent to which records are routed. Such a combination allows us to characterize operational sub-apportionment with respect to possible apportionments. We start with an aggregate analysis of record routability, but uncover that this approach provides a skewed perspective. Thus, we refine our approach through applying the forest analysis, evaluating record routability considering a record's location in the forest.

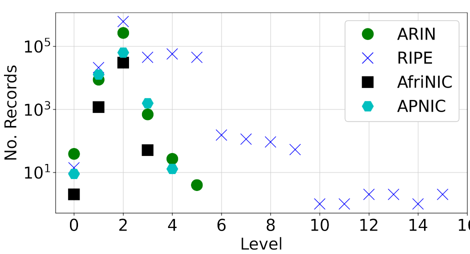
Aggregate Perspective. We begin with an aggregate evaluation of the extent to which WHOIS records are routed. For a given record, we quantify the portion of its address space that is routable (with any routes) using the HD Ratio (defined in Section 5.3 as an RIR address space utilization metric). We avoid simply quantifying the fraction of the record's address space which is routable



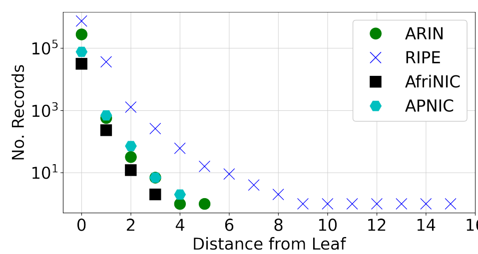
(a) The percent of records fully routable and unrouted with respect to their level in the tree.



(b) The percent of records fully routable and unrouted with respect to their distance from a leaf.



(c) The number of records in each level.



(d) The number of records from each leaf distance.

Fig. 8. In Figures 8a and 8b, we plot the percent of records that are completely routable (colored + textured) underneath the percent of records that are not routable at all (light gray), across levels in forest trees starting from the root or from the leaves, respectively. We note that in both figures there are more levels for RIPE, but we omit them for brevity as the number of records is small at those levels. In Figures 8c and 8d we show how many records exist at each level, from the root, and from the leaves, respectively.

as such metric skews towards small values (e.g., a /32 record routed across dozens of /48s will still exhibit a less than 1% routed address volume, as longer prefixes are exponentially smaller than shorter ones.)

We plot the HD Ratio distribution for all WHOIS records in Figure 7. We observe that this distribution is largely bimodal with the vast majority of records either being completely routed (85-99% across RIRs) or not routed at all (1-15% across RIRs), and only a small fraction of records exhibiting partial routing. However, we note that despite most records being routed, overall only a small fraction of the record address space is routed, as discussed in Section 4. We hypothesize that the structure of the WHOIS records can explain this discrepancy: if a parent record is routed, all of its children will be routed as well, resulting in a number of records covered by the same route(s). Thus looking across all WHOIS records provides a skewed perspective, and instead, we consider applying the forest analysis approach to better understand the routing characteristics of records.

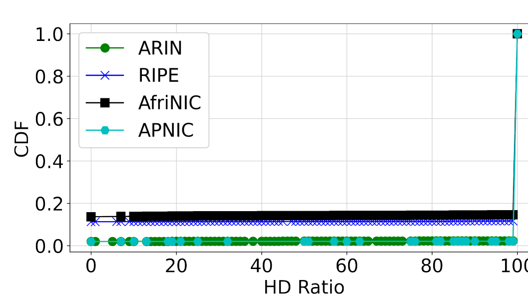


Fig. 7. CDF of the HD Ratios for all records. We see a bimodal distribution with nearly all records being completely routed or not routed at all.

Forest Analysis. Here, we evaluate the extent to which WHOIS records are routed, with respect to the record's position in the forest. We again quantify the portion of a record's address space that is routed using the HD Ratio. For each record in the forest, we identify its depth/level in the tree (with level 0 indicating a root record). In Figure 8a we plot the percentage of records that are routed completely or not at all. As with the aggregate perspective, we also observe at different levels that records primarily fall into a bimodal distribution with records either fully routed or completely unrouted. Only level 0 does not exhibit this bimodal property, because these are direct records owned by each RIR (and thus are unlikely to be fully routed). However, at deeper levels (e.g., level 2 or deeper), we observe that the distribution skews heavily towards fully routed records, whereas level 1 records are more evenly split between the two modes.

This observation, as well as the number of records per level (Figure 8c), confirms our hypothesis that a significant portion of large (level 1) records are fully routed, and all children records (if any) are thus also fully routed. For level 1 records that are partially routed, their children are also likely fully routed because these records are at finer granularities and are often designated for specific organizations and thus are either highly routed or not in use. We confirm this by analyzing record routability from the leaves upwards, as shown in Figure 8b. We see that for leaf records, the vast majority of records are fully routed (and leaves also account for the most records, as shown in Figure 8d). Thus, the most fully routed records are those deeper in the tree, particularly at leaves.

In this section, we identified the indicators of reachable IPv6 apportionment with respect to the forest of all possible apportionment. We found that apportionments in WHOIS records tend to be completely routable or not routable at all, and that these groups can be most meaningfully isolated by calculating their distance from a leaf node. This implies that, while routable apportionment is still a subset of that in WHOIS records, there are patterns in the ways providers choose this subset of their space. Regardless of a specific provider's suballocation structure and choices, routable area remains a subset that is most strongly identifiable by the most sub-apportioned part of their space.

9 CONTEXTUALIZING APPORTIONMENT WITH KNOWN ACTIVE ADDRESSES

While RIR WHOIS records give us the usable apportionments and routed prefixes give us a sense of potentially reachable apportionments, active addresses give us the strongest indication of *actually* utilized apportionment. Therefore, analysis of active addresses in the context of WHOIS records and routes is an important characterization of *active* apportionments. We thus take datasets of known IPv6 addresses (Section 4), and intersect these addresses with both RIR bulk WHOIS records as well as BGP routed prefixes to understand active apportionment.

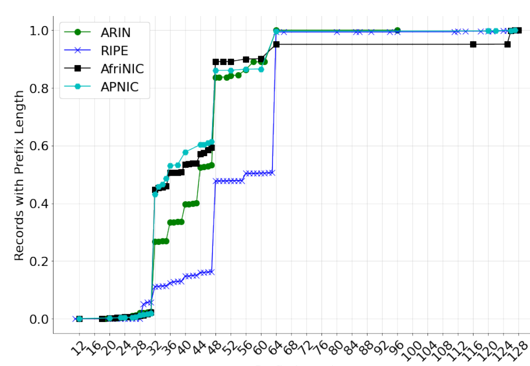


Fig. 9. CDF of prefix sizes for WHOIS records with any active addresses, for each RIR.

Active address datasets are fundamentally incomplete as it is infeasible to comprehensively survey all active IPv6 addresses, and such limitations will remain persistent in the future. However, we can analyze the addresses currently discoverable to identify potential correlations between record and route properties and the occurrence of active addresses. Given the incompleteness of our active address datasets though, we avoid analyzing the *number* of active addresses in an address space region, and rather consider the binary property of any active addresses associated with the region.

9.1 Active Addresses in WHOIS Records

Aggregate Perspectives. We map active addresses to the WHOIS record with the longest matching prefix, as such a record should be associated with the organization directly controlling the address. In total, 120,001 out of 1.178M records (10.2%) exhibit at least one active address with this mapping. Figure 9 depicts the CDFs of the prefix sizes for records with active addresses. For ARIN, AfriNIC, and APNIC, the most common records with activity are /32 and /48 prefixes. In contrast, /64 and /48 records are most common for RIPE, which we hypothesize are due to varying policies within providers in the RIPE geographic region, as several such organizations under RIPE's purvey contained a large number of /64 records (Section 5). Interestingly, for both RIPE and ARIN, the most common records overall are /64 and /56 prefixes, which does not align with their most active apportionments.

In our discussion of bulk records (Section 5.2), we noted an overwhelming amount of records are of size /48, /56, and /64. However, in Figure 9, we see this is not the case for records with active addresses. We see a much higher percentage of records with an active IP of size /32 (24.3% in ARIN, 5.3% in RIPE, 42.6% in APNIC, 41.3% in AfriNIC). From this discrepancy, we conclude that there is a fundamental difference between the intended use of smaller WHOIS records and their actual use. As discussed, per RIR policies (Section 5.3), records of size /48-/64 are intended for end-hosts. The act of creating a record for these prefixes indicates there was a specific intended use for it. However, if we are not able to observe any activity for prefixes of this size at large, we can conclude that their owners are not using them as expected, or at least in a way that we can observe.

We observe 177,138 addresses from our dataset that did not map to any WHOIS record. We find that all of these addresses are within the 2002::/16 prefix, associated with the 6to4 IPv6-to-IPv4 transition protocol, or other special use cases (as specified by IANA). Neither use cases indicate addresses that would be expected to map to WHOIS records or specific organization address ranges that would be in a record we include in this analysis. Thus, we do not observe discrepancies between active addresses and address space in WHOIS records.

Forest Analysis. We now evaluate via our forest analysis at each tree level separately and map active addresses to the WHOIS record at each level. In the left half of Table 4, we show the fraction of WHOIS records with active addresses, for each RIR and each tree level, from the root. Note Figures 8c and 8d showed the number of records per RIR and level. We observe that the distribution of active records varies greatly per RIR. However, overall we can generally look at level 1 to find records where the largest fraction is active, keeping in mind that moving further down the levels of the tree isolates records by specific providers with more frequent sub-apportionments. We hypothesize that RIPE's lower-level records are most active because of specific providers who have measurable addresses.

Dist.	ARIN ↓	APNIC ↓	RIPE ↓	AfriNIC ↓	ARIN ↑	APNIC ↑	RIPE ↑	AfriNIC ↑
0	23.1%	100%	100%	100%	3.31%	7.88 %	13.1%	3.03%
1	52.4 %	32.5%	48.8%	53.7 %	90.0 %	75.5%	18.9%	85.3%
2	1.78%	3.53%	13.7%	1.66%	84.4%	83.3%	91.2%	91.7%
3	48.9%	16.7%	15.7%	41.2%	100%	100%	100%	100%
4	40.7%	7.69%	5.94%	-	100%	100%	100%	-
5	75.0%	-	1.40%	-	100%	-	100%	-
6	-	-	55.5%	-	-	-	100%	-
7	-	-	57.0%	-	-	-	100%	-
8	-	-	53.3%	-	-	-	100%	-
9	-	-	9.43%	-	-	-	100%	-
10	-	-	100%	-	-	-	100%	-

Table 4. Percent of records at individual prefix's distance from root (↓) or leaf (↑), with at least 1 active IP. We highlight that when using the distance from a leaf one can reach groups of prefixes that are more and more active by our definition. We truncate the table at a distance of 10 due to low numbers of records past this.

On the other hand, in the right half of Table 4, we also show the fraction of WHOIS records with active addresses by the distance the record is from a leaf in the tree. As opposed to the left half of the table, we observe that every RIR reaches 100%. However, we point back to Figure 8d and note that this is not a significant number of records. We also note that the trend that we see with active IPs aligns with our finding in Section 8 that a distance of 0-2 levels from leaves is associated with the largest group of records that are routable. At these distances from leaves, we also find the most significant groups of records that all have at least 1 active address.

Once again, we conclude that using a record's distance from a leaf is a significant indicator of activity and that this characteristic can be used to find groups of active records.

9.2 Active Addresses in Routes

Next, we characterize the active addresses in BGP routes, focusing again on whether routes contain IP addresses or not, rather than the raw counts of IP addresses per route. We first map all active IP addresses to routes under the longest prefix match, mapping each address to the most specific route which would be used to reach it (also avoiding additionally counting parent routes). In total, 166,823 out of 223,210 routes (74.7%) exhibit at least one active address under this mapping.

Figure 10 shows the CDF of prefix lengths for routes with active addresses, for all routes as well as only those that are source routes. We use source routes to represent the boundaries of routable regions. We observe that active addresses are most likely in /48 routes, with /32 routes also common. These prevalences align with the common prefixes of routes being allocated and the routing recommendations described in Section 4 but highlight that in many cases smaller networks are advertised. We use this analysis to again conclude that active space is a subset of space apportioned by RIR data, supporting our usage of routes to contextualize records.

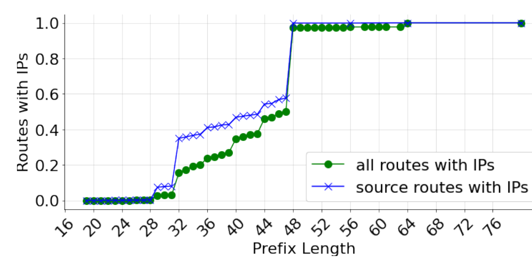


Fig. 10. Prefix Lengths for source routes and all routes with at least 1 active IP found. X Axis goes past 76, but after this the CDF is 1 so we omit.

We once again observe addresses from our dataset that do not map to any route. We observe 279,485 total of these addresses, 102,347 of which are within valid RIR records. This brings up questions of routability, however, we acknowledge that our routing dataset is not complete.

9.3 Combining All Datasets

Finally, we consider how WHOIS records, routes, and active IP addresses intersect. We observe that 95.4% of records with active IPs are completely routed, while 90.0% of records with no active IPs are completely routed. Thus, the routability of records is not strongly correlated with active addresses.

When looking at indicators of active addresses based on records and their routability, we observe a diversity of scenarios. To illustrate this heterogeneity, we present three case studies of distinct providers. We discuss the specifics of how each dataset varies across its characterization of apportionment per provider. We use these cases to show how our work finds meaningful apportionments of space relative to the RIR structure and routes. From these case studies, we show that the datasets are disjoint. We further emphasize the observation that IPv6 apportionment is used differently depending on the stakeholders in each area of space. Depending on the owner of the space, different approaches to leveraging various datasets is necessary to identify activity indicators.

Case Study 1: Comcast, 2001:558::/29. The allocation 2001:558::/29 is maintained by Comcast and is on the first level of ARIN's record structure with 1 level of suballocation below it. When inspecting the WHOIS record for this allocation, we observe 280 suballocations/assignments. All are direct children from this Comcast allocation, with record prefix sizes primarily being /48s with the exception of one /35. There are a total of 181,191 active IPs found across 108 records. The entire allocated prefix is routed with a single /29 route, but we observe 4,290 children routes as well, of which all but two are /48 prefixes. When mapping active addresses to routes, we see 172,529 IPs within the overall /29. However, there are only 4083 children routes containing 1,724 active IPs in total. Curiously, we see no overlap between Comcast's records of size /48 and the routes that we observe of size /48. There are 107 other Comcast WHOIS records of size /48 that contain active IPs, but the vast majority (174,253 IPs) are contained within Comcast's /29. This example highlights the inconsistency of sub-apportionment size between records and routes, and where active addresses appear in different sub-apportionment in each, but largely exist in the larger apportionment.

Case Study 2: Akamai, 2a02:26f0::/29. The allocation 2a02:26f0::/29 is on the first level of RIPE's structure with 1 level of suballocation beneath it. Within WHOIS records, we see 158 child records of size /48. Within the entirety of the space under the /29, we find 112,421 active addresses that span 130 WHOIS records. The allocation is routed in its entirety as 2a02:26f0::/29 and has 1,696 child routes, 1,691 of which are of length /48. We also find 1,240 IPs within the /29. However, in this case, active IPs are more concentrated in the smaller routes with 2.7% (3,004) of IPs concentrated in the 2a02:26f0:3400::/48 route. In fact, 97% of routes have at least 1 active IP. Here, we see a different view of active addresses across apportionments where activity exists primarily in sub-apportionments.

Case Study 3: Hostinger, 2a02:4780::/32. The allocation 2a02:4780::/32 is on the first level of RIPE's structure and has 1 level of suballocation beneath it. There are 21 child records, all of size /48. There are a total of 1,263,097 active addresses that we see within this allocation. Interestingly, these addresses are located in 21 of the records, including the parent /32. We do not observe routes for 2 records that contain active addresses. We also observe that some of the /48s have as many as 369,710 active addresses while some have as few as 1. The allocation is not routed in its entirety, rather there are 20 routes of size /48. Contrary to the former examples, the routed portions of this record crossover with some of the suballocations and assignments of this record. Of the 21 child records of size /48, 20 of them are the routes we observe. This example shows an overlap between sub-apportionment in routes and records exists, and this is a strong indicator of activity.

Summary. We observed fundamentally different behavior in active apportionment by defining it with active addresses both from a high level and specific to large providers. We observe that from the aggregate perspective, the size of active apportionments is different than that of all apportionments, and that large providers use their records and routes in observably different ways that point to distinct indicators of their active space. We, therefore, recommend that future work in IPv6 consider provider-specific properties in each area of the space, and emphasize that there is no "one size fits all" heuristic to IPv6 apportionment.

10 DISCUSSIONS AND RECOMMENDATIONS

We now summarize our key takeaways and synthesize recommendations from our work aimed at helping operationalize our insights into future IPv6 work:

- Providers vary greatly in their IPv6 apportionment behavior. This variation is directly visible in WHOIS records across providers and RIRs, such as with the inconsistent use of smaller records normally intended for end-sites (of prefix sizes /48-/64). The actual deployment of IPv6 is also heterogeneous as visible through routing data. For example, provider route sizes differ greatly based on whether they route sub-apportionments. This heterogeneity extends to indicators of active addresses in routes and records, where different providers exhibited varying indicators of which prefixes will contain active addresses. Together, these observations demonstrate that general heuristics on IPv6 apportionment, including those used in prior work, will fail to properly summarize real-world behavior, and instead per-provider analysis/evaluation is needed.

- Related, prior work [31] identified the challenges in determining a singular prefix or block for IP-based security mechanisms such as blocklisting or rate limiting. Our work suggests a dynamic granularity of apportionments for enforcement may alleviate such problems. A forest-based analysis approach showed patterns in provider behavior based on the levels of sub-apportionment, and thus may be more effectively leveraged for such applications.
- Prior work on IPv6 address generation operated under the assumption that address assignment patterns align within BGP route boundaries or at certain prefix sizes, and generate addresses based on inferences drawn from active addresses within those boundaries. Our findings highlight that route granularity is often too narrow, as a single provider's address space can contain numerous routes, and that providers exhibit wide variations in their prefixes, limiting the effectiveness of using static prefix sizes. Instead, address-generation methods can operate at the apportionment granularities, potentially providing meaningful boundaries with which to model address patterns.
- Delegation files provide a high-level overview of apportionment from the perspective of the RIR and largely do not show sub-apportionment structure. On the other hand, bulk records show the extent of sub-apportionment. Therefore, if it is possible to access them, we recommend using the bulk records for any use cases involving IPv6 apportionment, although delegation files may be appropriate for specific high-level use cases. However, when using this dataset, we caution that aggregate statistics may be skewed by the behavior of particular organizations.
- Although RIR policies are informative about how they intend to manage IPv6 apportionment, they are by no means ground truth, as assumed by some prior work. We recommend analyzing properties such as prefix sizes directly in real-world datasets in order to understand characteristics of apportionment in practice. Furthermore, characteristics of apportionment cannot be generalized across geographic regions. Therefore, we recommend analyzing per-RIR analysis.

We additionally refer back to the previous applications of IPv6 apportionment ideas in Section 3 and make the following recommendations for future IPv6 research:

- For active address generation, we suggest prioritizing generation within the address space regions at the intersection of routes and bulk WHOIS data with the forest-based approach. As described in Section 8, the distance-from-leaf measure can be used to more likely find groups of prefixes that will be fully reachable, but also provide significant coverage (i.e. there are enough prefixes to meaningfully analyze).
- For grouping addresses together, the best apportionment depends on the level of detail desired in the analysis of addressing patterns. For a provider-level view, we recommend using the delegation files. As found in Section 5.4, the delegation files primarily contain records that are sub-apportioned one level from the RIR's original allocations. Therefore, the delegation files provide the prefix size per provider but do not include the prefix sizes of the specific services in smaller records listed in the bulk WHOIS data.
- If it is not possible to update a prefix size parameter in an algorithm for target generation or address analysis, then it is important to look at the WHOIS datasets for the best prefix size to choose for the parameter. As observed in Section 5, the broad RIR policies do not always reflect what is actually occurring and often organizations use prefix sizes that do not conform to policies/standards.

11 CONCLUSION

In this work, we conducted an analysis of the apportionment of the IPv6 address space from the ground-up, evaluating multiple distinct data sources both individually and in tandem. Specifically, we considered how the IPv6 address space was distributed amongst different providers and organizations from RIR WHOIS data, and how these organizations use their address space both through deploying BGP routes to make addresses potentially reachable and through exhibiting active hosts on those addresses. Through our investigation, we identified that the perspectives offered by these varying datasets paint fundamentally different pictures of IPv6 apportionment in terms of both size and structural properties, shedding light on the behavior of providers and organizations when using IPv6. We identified that simplifying assumptions or heuristics on apportionment behavior, including those used by prior work (especially in terms of subnet sizes and prefix lengths), are rarely applicable, as providers exhibit a diversity of operational behavior. Instead, provider-specific characterizations are needed. To that end, we leveraged a forest-based approach for evaluating these IPv6 datasets, which avoids the skewed perspective offered by aggregate analyses. Ultimately, our study provides a deeper understanding about IPv6 deployments in practice, lessons for using these datasets, as well as insights for operationalizing their analyses in the future.

Moving forward, there are salient directions to investigate further. One direction is in investigating the address assignment patterns within the individual network provider prefixes that we identified in this work based on WHOIS records and structure. Entropy analysis on active IP addresses per provider could identify clusters of common assignment behaviors, which would provide deeper insights into meaningful apportionment boundaries to operationalize. Another direction is in characterizing IPv6 apportionment over time. Such longitudinal evaluation would allow us to better understand the dynamics of IPv6 growth and utilization, including how structural dimensions evolve temporally. Such understanding could also afford predictions of future activity in nascent regions of the IPv6 address space. A third proposed direction is in empirically evaluating the operationalization of our apportionment boundaries, as discussed earlier. For example, future work can evaluate the effectiveness of different methods for IPv6 address generation within apportionment boundaries for Internet scanning purposes. Together, such directions will provide more effective approaches for understanding and improving IPv6 measurements.

12 ACKNOWLEDGEMENTS

This work was supported in part by a Georgia Tech Research Institute (GTRI) Independent Research and Development (IRAD) grant, and by a National Science Foundation (NSF) Graduate Research Fellowship (GRFP) under Grant No. DGE-2039655. The authors would like to thank Adonis Bovell for his feedback and assistance with collecting data.

REFERENCES

[1] AFRINIC. Afrinic public ftp server, 2022. https://ftp.afrinic.net/stats/afrinic/.
[2] Afrinic bulk whois data. https://afrinic.net/support/whois/bulk.
[3] Policy to change the ipv6 hd ratio from 0.8 to 0.94. https://afrinic.net/policy-to-change-the-ipv6-hd-ratio-from-0-8-to-0-94-afpub-2007-v6-002#:~:text=The%20HD%2DRatio%20presently%20defined,the%20HD%20ratio%20of%200.80.
[4] APNIC. Apnic internet number resource policies, 2021. https://www.apnic.net/community/policy/resources,https://www.apnic.net/community/policy/resources.
[5] APNIC. Apnic public ftp server, 2022. https://ftp.apnic.net/stats/apnic/.
[6] Apnic bulk whois data. https://www.apnic.net/manage-ip/using-whois/bulk-access/.
[7] ARIN. Number resource policy manual, 2021. https://www.arin.net/participate/policy/nrpm/.
[8] ARIN. Arin public ftp server, 2022. https://ftp.arin.net/pub/stats/arin/.
[9] Arin bulk whois data. https://www.arin.net/reference/research/bulkwhois/.
[10] Ark ipv4 routed /24 topology. https://catalog.caida.org/details/dataset/ark_ipv4_traceroute.
[11] Beverly, R., Durairajan, R., Plonka, D., and Rohrer, J. P. In the IP of the beholder: Strategies for active ipv6 topology discovery. In *Proceedings of the Internet Measurement Conference 2018, IMC 2018, Boston, MA, USA, October 31 - November 02, 2018* (2018), ACM, pp. 308–321.

[12] The caida ucsd ipv6 dns names dataset - <12/19/2021>. https://www.caida.org/catalog/datasets/ipv6_dnsnames_dataset/.

[13] The caida ucsd ipv6 topology dataset - <12/19/2021>. https://www.caida.org/catalog/datasets/ipv6_allpref_topology_dataset/.

[14] Cho, K., Mitsuya, K., and Kato, A. Traffic data repository at the wide project.

[15] Czyz, J., Allman, M., Zhang, J., Iekel-Johnson, S., Osterweil, E., and Bailey, M. Measuring ipv6 adoption. *SIGCOMM Comput. Commun. Rev. 44*, 4 (aug 2014), 87–98.

[16] Durumeric, Z., Adrian, D., Mirian, A., Bailey, M., and Halderman, J. A. A search engine backed by internet-wide scanning. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security* (New York, NY, USA, 2015), CCS '15, Association for Computing Machinery, p. 542–553.

[17] Durumeric, Z., Wustrow, E., and Halderman, J. A. Zmap: Fast internet-wide scanning and its security applications. In *Proceedings of the 22nd USENIX Conference on Security* (USA, 2013), SEC'13, USENIX Association, p. 605–620.

[18] Farsight. Dnsdb, 2021. https://www.farsightsecurity.com/solutions/dnsdb/.

[19] Foremski, P., Plonka, D., and Berger, A. Entropy/ip: Uncovering structure in ipv6 addresses. In *Proceedings of the 2016 Internet Measurement Conference* (New York, NY, USA, 2016), IMC '16, Association for Computing Machinery, p. 167–181.

[20] Gasser, O., Scheitle, Q., Foremski, P., Lone, Q., Korczyński, M., Strowes, S. D., Hendriks, L., and Carle, G. Clusters in the expanse: Understanding and unbiasing ipv6 hitlists. In *Proceedings of the Internet Measurement Conference 2018* (New York, NY, USA, 2018), IMC '18, Association for Computing Machinery, p. 364–378.

[21] Gasser, O., Scheitle, Q., Gebhard, S., and Carle, G. Scanning the ipv6 internet: Towards a comprehensive hitlist. In *Traffic Monitoring and Analysis - 8th International Workshop, TMA 2016, Louvain la Neuve, Belgium, April 7-8, 2016* (2016), A. Botta, R. Sadre, and F. E. Bustamante, Eds., IFIP.

[22] Google. Ipv6 statistics, 2021. https://www.google.com/intl/en/ipv6/statistics.html.

[23] Heidemann, J., Pradkin, Y., Govindan, R., Papadopoulos, C., Bartlett, G., and Bannister, J. Census and survey of the visible internet. In *Proceedings of the ACM Internet Measurement Conference* (Vouliagmeni, Greece, Oct. 2008), ACM, pp. 169–182.

[24] Ark ipv4 routed /24 topology. https://www.iana.org/numbers.

[25] Izhikevich, L., Akiwate, G., Berger, B., Drakontaidis, S., Ascheman, A., Pearce, P., Adrian, D., and Durumeric, Z. Zdns: A fast dns toolkit for internet measurement. In *Proceedings of the 22nd ACM Internet Measurement Conference* (New York, NY, USA, 2022), IMC '22, Association for Computing Machinery, p. 33–43.

[26] Izhikevich, L., Teixeira, R., and Durumeric, Z. Predicting ipv4 services across all ports. In *Proceedings of the ACM SIGCOMM 2022 Conference* (New York, NY, USA, 2022), SIGCOMM '22, Association for Computing Machinery, p. 503–515.

[27] Karpilovsky, E., Gerber, A., Pei, D., Rexford, J., and Shaikh, A. Quantifying the extent of ipv6 deployment. In *Passive and Active Measurement Conference (PAM)* (04 2009), pp. 13–22.

[28] LACNIC. Policy manual, 2020. https://www.lacnic.net/680/2/lacnic/.

[29] LACNIC. Lacnic public ftp server, 2022. https://ftp.lacnic.net/pub/stats/lacnic/.

[30] Lacnic bulk whois data. https://www.lacnic.net/2472/2/lacnic/request-bulk-whois-access.

[31] Li, F., and Freeman, D. Towards a user-level understanding of ipv6 behavior. In *Proceedings of the ACM Internet Measurement Conference* (New York, NY, USA, 2020), IMC '20, Association for Computing Machinery, p. 428–442.

[32] The majestic million. https://majestic.com/reports/majestic-million.

[33] Murdock, A., Li, F., Bramsen, P., Durumeric, Z., and Paxson, V. Target generation for internet-wide ipv6 scanning. In *Proceedings of the 2017 Internet Measurement Conference* (New York, NY, USA, 2017), IMC '17, Association for Computing Machinery, p. 242–253.

[34] Nemmi, E. N., Sassi, F., La Morgia, M., Testart, C., Mei, A., and Dainotti, A. The parallel lives of autonomous systems: Asn allocations vs. bgp. In *Proceedings of the 21st ACM Internet Measurement Conference* (New York, NY, USA, 2021), IMC '21, Association for Computing Machinery, p. 593–611.

[35] of Oregon, U. Routeviews project, 2021. http://www.routeviews.org/routeviews.

[36] Padmanabhan, R., Rula, J. P., Richter, P., Strowes, S. D., and Dainotti, A. Dynamips: Analyzing address assignment practices in ipv4 and ipv6. In *Proceedings of the 16th International Conference on Emerging Networking EXperiments and Technologies* (New York, NY, USA, 2020), CoNEXT '20, Association for Computing Machinery, p. 55–70.

[37] Plonka, D., and Berger, A. Temporal and spatial classification of active ipv6 addresses. In *Proceedings of the 2015 Internet Measurement Conference* (New York, NY, USA, 2015), IMC '15, Association for Computing Machinery, p. 509–522.

[38] Rapid7 forward dns. https://opendata.rapid7.com/sonar.fdns_v2/.

[39] Rfc 1715. https://datatracker.ietf.org/doc/html/rfc1715.

[40] Rfc 3194. https://datatracker.ietf.org/doc/html/rfc3194.

[41] Rfc 4692. https://datatracker.ietf.org/doc/html/rfc4692.
[42] RIPE. Ripe public ftp server, 2022. https://ftp.ripe.net/pub/stats/ripencc/.
[43] Ripe atlas dataset. https://data-store.ripe.net/datasets/atlas-daily-dumps/.
[44] RIPE-NCC. Ipv6 address allocation and assignment policy, 2020. https://www.ripe.net/publications/docs/ripe-738.
[45] RIPE-NCC. Routing information service (ris, 2022. https://www.ripe.net/analyse/internet-measurements/routing-information-service-ris.
[46] Best current operational practice for operators: Ipv6 prefix assignment for end-users - persistent vs non-persistent, and what size to choose. https://www.ripe.net/publications/docs/ripe-690.
[47] Rohrer, J. P., LaFever, B., and Beverly, R. Empirical study of router ipv6 interface address distributions. *IEEE Internet Computing 20*, 4 (jul 2016), 36–45.
[48] Rye, E., Beverly, R., and Claffy, K. C. Follow the scent: Defeating ipv6 prefix rotation privacy. In *Proceedings of the 21st ACM Internet Measurement Conference* (New York, NY, USA, 2021), IMC '21, Association for Computing Machinery, p. 739–752.
[49] Sediqi, K. Z., Prehn, L., and Gasser, O. Hyper-specific prefixes. *ACM SIGCOMM Computer Communication Review 52*, 2 (apr 2022), 20–34.
[50] Strowes, S. Visibility of ipv4 and ipv6 prefix lengths in 2019, 2019. https://labs.ripe.net/author/stephen_strowes/visibility-of-ipv4-and-ipv6-prefix-lengths-in-2019/.
[51] Ullrich, J., Kieseberg, P., Krombholz, K., and Weippl, E. On Reconnaissance with IPv6: A Pattern-Based Scanning Approach. In *2015 10th International Conference on Availability, Reliability and Security* (Toulouse, France, Aug. 2015), IEEE, pp. 186–192.
[52] Cisco umbrella popularity list. http://s3-us-west-1.amazonaws.com/umbrella-static/index.html.
[53] Wei, J. Why is a /48 the recommended minimum prefix size for routing?, 2020. https://blog.apnic.net/2020/06/01/why-is-a-48-the-recommended-minimum-prefix-size-for-routing/.
[54] Zirngibl, J., Steger, L., Sattler, P., Gasser, O., and Carle, G. Rusty clusters? dusting an ipv6 research foundation. In *Proceedings of the 22nd ACM Internet Measurement Conference* (New York, NY, USA, 2022), IMC '22, Association for Computing Machinery, p. n/a.

A ETHICS

We do not believe this work raises ethical issues. In order to grow out our active IPv6 address dataset, we use ZDNS [25] to resolve names we find in publicly accessible domain name data sources, including Censys [16], CT logs, Majestic 1 Million [32], and Cisco Umbrella [52]. These are our only active measurements, and our remaining datasets are all public.

Received October 2022; revised December 2022; accepted January 2023