DynamIPs: Analyzing address assignment practices in IPv4 and IPv6

Ramakrishna Padmanabhan
CAIDA, UC San Diego
ramapad@caida.org

John P. Rula
Akamai
jrula@akamai.com

Philipp Richter
Akamai / MIT
prichter@akamai.com

Stephen D. Strowes
RIPE NCC
sds@ripe.net

Alberto Dainotti
CAIDA, UC San Diego
alberto@caida.org

ABSTRACT

IP addresses are commonly used to identify hosts or properties of hosts. The address assigned to a host may change, however, and the extent to which these changes occur in time as well as in the address space is currently unknown, especially in IPv6.

In this work, we take a first step towards understanding the dynamics of IPv6 address assignments in various networks around the world and how they relate to IPv4 dynamics. We present fine-grained observations of dynamics using data collected from over 3,000 RIPE Atlas probes in dual-stack networks. RIPE Atlas probes in these networks report both their IPv4 and their IPv6 address, allowing us to track changes over time and in the address space. To corroborate and extend our findings, we also use a dataset containing 32.7 billion IPv4 and IPv6 address associations observed by a major CDN. Our investigation of temporal dynamics with these datasets shows that IPv6 assignments have longer durations than IPv4 assignments—often remaining stable for months—thereby allowing the possibility of long-term fingerprinting of IPv6 subscribers. Our analysis of spatial dynamics reveals IPv6 address-assignment patterns that shed light on the size of the address pools network operators use in domestic networks, and provides preliminary results on the size of the prefixes delegated to home networks. Our observations can benefit many applications, including host reputation systems, active probing methods, and mechanisms for privacy preservation.

CCS CONCEPTS

• **Networks** → Network measurement;

ACM Reference Format:
Ramakrishna Padmanabhan, John P. Rula, Philipp Richter, Stephen D. Strowes, and Alberto Dainotti. 2020. DynamIPs: Analyzing address assignment practices in IPv4 and IPv6. In *The 16th International Conference on emerging Networking EXperiments and Technologies (CoNEXT '20), December 1–4, 2020, Barcelona, Spain.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3386367.3431314

CoNEXT '20, December 1–4, 2020, Barcelona, Spain

ACM ISBN 978-1-4503-7948-9/20/12…$15.00
https://doi.org/10.1145/3386367.3431314

1 INTRODUCTION

Originally, the Internet architecture called for IP addresses to identify network device interfaces, effectively allowing the mapping of IP addresses to individual hosts. As a result, numerous academic projects and commercial products that make inferences based on IP addresses have emerged. Prime examples include geolocation databases to map individual IP addresses to geographic locations, as well as host reputation databases, e.g., to enable blocking of traffic from known malicious hosts. Some studies even use IP addresses to track hosts over time; for example to estimate the host-count in peer-to-peer networks [7] and botnets [39], or the number of open resolvers in the Internet [10, 53]. IP addresses also have been used as a proxy for CPE (Customer Premises Equipment) to determine when they experience outages [35, 54]. In many of these applications, there exists an expectation that a host's IP address will persist for sufficient time.

However, the extent to which this expectation is valid is unknown, particularly for dual-stacked Internet hosts. As networks transition from IPv4 to IPv6, many hosts on the Internet are dual-stacked, i.e., have both an IPv4 and an IPv6 address, and addressing strategies can vary significantly between the two protocols. In the case of IPv4, fueled by address space scarcity and fragmentation [44], addresses are assigned to end users in numerous ways, including dynamic address assignment with varying lease times and assignment policies, and Carrier-Grade NAT to multiplex more users behind fewer public IPv4 addresses. In the case of IPv6, operators have myriad ways to leverage the vast IPv6 space to subnet their own network and to delegate prefixes of various lengths to their subscribers. A typical subscriber's CPE device (the router inside a subscriber's home) receives entire IPv6 prefix delegations, which the CPE can further subdelegate to connected devices.

The co-existence of two addressing protocols, paired with increasing complexity of address assignment practices in both protocols, make it largely infeasible to answer even simple questions such as "how long can I expect an IPv4/IPv6 address to be assigned to a host?" or "is there an IPv4/IPv6 prefix size that can identify an individual host or groups of hosts?". This uncertainty is a major blow both for host reputation systems that aim to attribute malicious activity to individual hosts (identified using IP addresses), as well as for measurement approaches that need stable addresses to carry out active probing, be it for outage detection, or to identify and track nascent Internet vulnerabilities.

To shed light on addressing in today's Internet, we present a broad and detailed study of some of the fundamental properties of IP

address assignment[1] in the Internet, focusing on temporal aspects, i.e., how long IP addresses are assigned, as well as spatial aspects, i.e., where do addresses move upon reassignment. We derive actionable insights from our measurements that may help researchers and practitioners when developing IP-based systems. While our primary focus is addressing in the IPv6 Internet, we present statistics for both protocols, and highlight their interactions. Our contributions are as follows:

IPv4 and IPv6 addressing in time: We provide a longitudinal and detailed assessment of how IPv4 addresses and IPv6 /64 prefixes are assigned in today's Internet. Leveraging a 6-year dataset gathered from RIPE Atlas probes, we study the duration of address assignments in IPv4 and IPv6 on over 3,000 dual-stack probes. *We find that IPv6 prefixes delegated to residential subscribers can remain stable for months, permitting long-term use of IPv6 prefixes to identify individual subscribers (at the CPE granularity), even if subscribers' devices are using privacy addresses.*

IPv4-IPv6 interplay: Using a dataset from a major CDN capturing 32.7 billion IPv4-IPv6 address associations from dual-stacked hosts, we investigate if and to what extent it is possible to associate individual IPv4 prefixes with their IPv6 counterparts, and how stable such associations are. *We corroborate our findings on assignment durations on a broad scale and also show how different IPv4 multiplexing mechanisms affect address associations.*

IPv6 addressing in space: We investigate spatial properties of address assignments, i.e., how far apart subsequent assignments to the same host are in IPv4 and IPv6. We find that IPv6 /64 prefix assignments are more spatially stable over time than in IPv4 and that subsequent assignments of IPv6 prefixes typically do not switch to a different routed BGP prefix, as opposed to IPv4. Leveraging our insights, we proceed to derive actionable knowledge about the emerging IPv6 space. *In particular, we identify IPv6 prefix boundaries that isolate groups of individual hosts, i.e., address pools. Further, we develop a technique to infer IPv6 prefix lengths that identify individual subscribers for some ISPs.*

Our work presents a solid first step towards deriving actionable knowledge about IPv4 and IPv6 addressing in today's Internet. Our findings add a vital component to support nascent approaches that rely on IP address information, since several applications would benefit from understanding the temporal and spatial stability of IP address assignments. Host reputation services would benefit from knowing how long to associate an address with a particular host and which prefixes the host can eventually move to. Host-tracking applications that use IP addresses to count phenomena (such as the number of botnet bots [39]) can reason more effectively about errors. Measurement studies that use IPv6 "hitlists" typically rely on spatial structure in addressing to scope the workload to a feasible number of targets; with knowledge of temporal and spatial aspects, hitlists may be able to reduce the scope of target addresses within particular networks. Further, an assessment of the temporal and spatial stability of residential IPv4 and IPv6 addresses can inform the debate on whether and when to consider IP addresses as PII. The majority of our findings are derived from publicly available data and our processed findings are available to the research community [40].

[1]We use the term "address assignment" generically to describe the assignment of an IPv4 address or the delegation of an IPv6 prefix.

This paper is structured as follows: We introduce background on addressing and related work in Section 2. We study IP address assignments over time in Section 3 and study IPv4/IPv6 associations in Section 4. We proceed to study IP address assignments in space in Section 5. We next discuss the applications of our findings in Section 6 and conclude in Section 7.

2 BACKGROUND AND RELATED WORK

2.1 IP address assignment strategies

With increasing adoption of IPv6 [8, 12, 64], many Internet hosts today are dual-stacked, and possess both an IPv4 and an IPv6 address. However, as the successor of IPv4, IPv6 has some fundamental differences in design—such as the use of 128 bits for addressing—that lead to differences in the structure and deployment of IPv4 and IPv6 addresses in residential networks. A typical residential network consists of a CPE, or "home router", and several devices within the home that connect to the CPE. The CPE in turn connects to the upstream ISP and routes packets between residential devices and the Internet. To enable residential devices to send Internet packets, the ISP provides configuration information to the CPE, including which addresses to use, and which DNS servers to contact. While a CPE may typically receive only a single IPv4 address, it can receive a prefix as large as a /48 in IPv6.

IPv4 residential deployments. Domestic ISPs commonly use DHCP [13] or RADIUS [47] to assign IPv4 addresses. Each CPE receives a single globally routable address that is shared by all devices in the customer's local network. The CPE operates as a DHCP server for hosts inside the home network, issuing addresses drawn from address ranges for private networks [43]. The CPE also operates NAT, to allow traffic to pass from private to public address space and back.

An extension to this model is Carrier-Grade NAT (CGNAT), where the CPE is assigned a private address from a different private range (e.g., [62]) and shares an upstream NAT function with multiple households [28, 46]. In cellular networks, devices are often assigned addresses from private ranges and the operator operates CGNAT on behalf of many subscribers to reach the public Internet.

IPv6 residential deployments. IPv6 addresses have been designed to have a network component and a host component, each 64-bits in length. ISPs are often allocated large address blocks, from which they can devise their own addressing architecture. Thus, IPv6 provides additional flexibility in assignment practices to operators.

In a residential IPv6 network, the CPE typically uses DHCPv6 to request an IPv6 *delegated prefix* from the ISP, and the assignment of a /56 address block to customer premises is a common recommendation [60]. The CPE itself is then free to suballocate /64s from that block and advertise them on the local home network. Hosts—unless configured statically or to use DHCPv6—specify the host component of the address autonomously. This component historically was autonomously generated and stable [56], though modern systems regularly generate a new host component for privacy reasons [18, 32]. Thus, the addresses on devices are globally routable, without address translation to reach the Internet. Note that the CPE is also likely to be assigned a prefix for the point-to-point subnet between the CPE and the ISP (the WAN prefix) [51]. The WAN prefix can be different from the LAN prefix obtained

via DHCPv6 [60]; our focus in this work is upon the dynamics of the LAN prefix. These address components introduce additional nuances in the study of IPv6 addressing characteristics: the delegated prefix managed by the CPE, the sub-prefixes assigned to local networks by the CPE, and the host components of addresses, may all change and at different times.

In this work, we focus upon the dynamics of the 64-bit "network" component of IPv6 addresses. Prior work has shown that the 64-bit host part of IPv6 addresses is often ephemeral [36] due to the widespread use of "privacy addresses"[2]. However, the 64-bit network component in addresses can also uniquely identify an individual subscriber since all the subscriber's devices may have addresses assigned from the same /64 prefix. Consequently, tracking /64 prefixes can allow tracking subscribers, leading to various applications and also to privacy implications, that we describe in Section 2.3.

2.2 Reasons address assignments change

To accommodate varying demand for IP addresses as subscribers connect and disconnect, ISPs typically assign addresses dynamically. ISPs have *pools* of addresses or prefixes from which addresses are assigned to subscribers by a DHCP/RADIUS server that is responsible for these pools. Typically, addresses in the same pool are related to each other, in network topology and geography, to facilitate the routing of Internet packets destined to these addresses.

While IP addresses can in theory be assigned to CPEs indefinitely, there are several factors that could lead to a change in the assigned address (or prefix, in the case of IPv6). Maier et al. describe why IPv4 addresses assigned via RADIUS can change [29] and Padmanabhan et al. provide a general overview of the reasons dynamically assigned IPv4 addresses change [34]. We summarize these causes and provide additional context for IPv6 prefix changes below.

Periodic changes. In IPv4, the scarcity of addresses forces ISPs to conserve addresses, leading to address assignment policies designed to reclaim addresses that are no longer in active use. A mechanism that helps with address conservation is a *lease*: addresses assigned to CPEs are only valid for the duration indicated in these leases. Once an address's lease has expired, it is reclaimed into the address pool and is available for assignment. The DHCP protocol allows the CPE to renew the lease before lease expiry [13], so that the CPE may continue to retain its address for multiple lease durations. However, addresses assigned via RADIUS typically change after the configured *SessionTimeout* (the equivalent of the DHCP lease time). Prior work has found well-defined periodic address changes in IPv4—for instance, after a period of 24 hours for a major European ISP [29] and periods of 24 hours, 36 hours, 1 week, 2 weeks etc. in several European and Asian ISPs [34].

IPv6 has no shortage of addresses but prefix assignments nevertheless change periodically in some ISPs, as we show in Section 3. Many networks use similar infrastructure for IPv4 and IPv6 and may employ similar mental models; they may therefore have similar policies for IPv6 assignment as they do for IPv4. Further, the privacy risks that arise from maintaining a persistent identifier for a CPE may lead ISPs to change assigned prefixes periodically.

[2]"Privacy addresses" are those whose 64-bit host components are randomly generated by user devices using SLAAC with privacy extensions [32].

Changes due to outages. In both IPv4 and IPv6, an outage affecting the upstream ISP's server/router that is responsible for routing packets to subscriber's addresses can result in loss of state about the addresses assigned to each subscriber. Such outages that affect ISP's infrastructure devices can result in new address assignments.

Outages of subscriber equipment can also result in address changes. Although DHCP offers clients the opportunity to renew leases and keep the same assignment, outages affecting the CPE that last longer than the lease duration may prevent the CPE from renewing the lease [34]. In some networks, even very short CPE outages or reboots can result in assignment changes [29, 34]; such changes typically occur when the server that assigns addresses does not maintain state about previously assigned addresses, as is typical in ISPs assigning addresses using RADIUS.

Administrative changes. One of the design goals of dynamic address assignment is to facilitate network reconfiguration. Thus, assignment changes can occur due to network renumbering by the ISP. Several factors could necessitate renumbering, including network restructuring, IP address acquisitions/losses during mergers, and changes in address pools to balance supply and demand.

2.3 Related work

Prior work has mostly focused upon temporal aspects of IPv4 address assignment using several approaches. UDMap used Hotmail user login traces to study dynamic address assignment [63] and Casado et al. tracked CDN clients using HTTP cookies [6]. Moura et al. used Zmap to ping the entire address space of large ISPs and identified session durations by observing continuous periods of responsiveness [30]. However, these studies' findings are inconsistent: some found that the majority of IPv4 addresses are associated with the same hosts for months [2, 6] while others reported that addresses can change after hours [29, 30, 63]. While these inconsistencies may arise due to a variety of causes—including differences in coverage, changes in assignment policies, and measurement artifacts—our study sheds light on one potential cause of inconsistency: non-dual-stack and dual-stack hosts in the *same* network can have different address assignment patterns (Section 3.2).

The most direct prior work to this paper studied both temporal and spatial characteristics of IPv4 address assignments throughout 2015 [34]. The authors showed that customer IPv4 addresses in some cases change frequently and regularly, typically in European and Asian networks. Due to the nature of the dataset used, dual-stacked RIPE Atlas probes were not included in that prior work. The dataset we leverage for our study not only sheds light on IPv6 practices, but allows us to study assignment patterns on dual-stacked hosts.

To the best of our knowledge, we are the first to conduct a detailed analysis of the spatial and temporal aspects of IPv6 /64 prefix assignment. Plonka and Berger studied characteristics of IPv6 addresses observed at a CDN over various time frames in 2014 and 2015 [36]. They identified substantial differences in addressing patterns across networks, and also provided results indicating that many /64s are short-lived: while they observed 90% of IPv6 /64s as active for at least a contiguous run of three days, fewer than 30% were stable for six months. In our work, we take full advantage of the RIPE Atlas platform to identify not only *when* network address

changes take place, but also *within which address ranges*, thus providing novel and up-to-date insights into these practices. Various applications and research areas can be impacted by the findings we present. We introduce them in the remainder of this section and discuss implications of our work for these approaches in Section 6.

Host reputation. Content and service providers incorporate multiple signals in order to measure host reputation and acquire threat intelligence to protect their services [5, 27, 42]. Reputation scoring tries to flag hosts or networks thought to be engaged in malicious activities such as email spam [14], amplification attacks [9], phishing, participating in botnets, and so forth. Reputation monitoring systems at the AS-level have also been proposed [24].

Blocklists are a common approach that operators use to temporarily filter traffic from bad actors but this approach can result in collateral damage to legitimate users [41]. Our findings on the durations for which IPv4 addresses are assigned to hosts can inform how long addresses can continue to remain on blocklists without causing collateral damage. In IPv6, it is not sufficient to block individual addresses (recall, a host can generate those independently from the network); instead, there is a tradeoff between blocking a short prefix for a long time as opposed to a longer prefix for a short time [26]. Blocking at the granularity of a /64 is more typical [38]. We know, however, that an individual subscriber can be delegated a prefix shorter than a /64 [60], potentially allowing evasion. Yet, the absence of per-network ground-truth data prohibits more comprehensive blocking. The results we present here provide insight on common practices at ISPs.

Tracking and Anonymity. Applications that track the number of users in a system can use our results and datasets to reason about the potential to "double-count" the same host multiple times due to dynamic reassignment and access over both IPv4 and IPv6 [52].

Stable interface identifiers (IIDs, the "host" component in an IPv6 address) based on link-layer addresses are no longer recommended [20]. However, many devices continue to use IID addressing based on the link-layer address (such addresses are called EUI-64 addresses), as observed in various studies [3, 17]. Our results on active address ranges in use indicate that these devices will be trackable across network address changes. Plonka and Berger investigated structure in IPv6 address sets to assist the sharing of large IPv6 datasets with reasonable confidence of maintaining privacy [37].

Identifying stable addresses for active probing. The IPv6 address space is several orders of magnitude larger than IPv4's. Consequently, most IPv6 addresses will not be in active use and, with the exception of "aliased prefixes" [17, 31], will not respond to active measurement techniques. Viable IPv6 targets for active measurement must therefore be curated and several studies have investigated how to generate lists of such targets, from e.g., DNS [15, 59] and DNSSEC infrastructure [4], the Bitcoin network, traceroute hops, and from a combination of techniques [17]. Similar work by Beverly et al. constructed measurement hitlists and additionally attempted to optimize topology discovery with Yarrp6 to reduce redundant traceroutes [3]. Rye and Beverly used Edgy to study IPv6 subnets allocated to the links between the provider networks and CPEs [51]. They identified daily lifetimes of these subnets in some ISPs, and also widespread use of stable EUI-64 IIDs. While their focus was upon the WAN prefix (Section 2.1), we study the LAN prefix, which is the subnet allocated to the subscriber premises (and therefore to the endpoints within the home network).

Hitlist curation implies that addresses are removed when no longer active. In this paper, we enumerate how often subscriber network allocations can change, implying many viable targets in residential or cellular networks will move to a new network address. Our results on active address structure in given networks can augment hitlists: identifying such structure offers a tractable set of network addresses to scan.

Target generation for active scanning. An extension of target curation is to use them to identify additional new targets. Knowledge of structure in IPv6 addressing has long been discussed [19]. Ullrich et al. used pattern-based scanning to locate targets within a /64 [58]. Foremski et al. presented Entropy/IP [16], which uses a machine learning technique trained on sets of full addresses (not just IIDs) to model the addressing schemes in use and generate new addresses for probing. Murdock et al. presented 6Gen [31] which does not try to learn structure, and instead aims to find dense regions in a seed list of IPv6 addresses to then generate neighboring addresses. These techniques rely on address sets of sufficient volume to identify structure and could be augmented with our findings.

2.4 Ethical Considerations

The RIPE Atlas data that we use is public and available for responsible use. We discuss the RIPE Atlas data in Section 3.1. All RIPE Atlas probes participate in the measurements that generate the data we use as the basis for this study. The CDN data described in Section 4.1 is captured on an IPv4 /24 and IPv6 /64 granularity, does not contain device identifiers, and complies with the CDN's own data privacy practices. Some of our results impact aspects of network security. Although one of the core interests of the research community is active network measurement, we did not use this data to perform active measurements as part of this study. It is of critical importance that providers are aware of common practices in IPv6 deployments, and also that such practices are taken into account as network deployments and addressing patterns evolve.

3 ADDRESSES IN TIME

We use the RIPE Atlas "IP echo" measurement dataset [48, 49] to begin our investigation of IPv4 and IPv6 assignment durations. In 2016, Padmanabhan et al. demonstrated the value in using the "connection logs" dataset collected from RIPE Atlas probes to study IPv4 address changes in non-dual-stack probes. In this work, by using the "IP echo" dataset, we are able to study IPv6 assignment changes and perform comparisons with IPv4. Further, we show that there are differences in *IPv4 assignment patterns* between dual-stack and non-dual-stack networks, providing a key update to state-of-the-art knowledge of IPv4 deployment.

3.1 RIPE Atlas "IP echo" dataset

Every hour, all RIPE Atlas probes automatically run IP echo measurements for both address types: at each iteration, a RIPE Atlas probe performs an HTTP GET request to an HTTP server operated by RIPE, which in turn returns in the response header a field with the key *X-Client-IP* and the value set to the client's IP address as visible to the HTTP server. These measurements thus regularly

"echo" back to the probe the public routable address that successfully opened a TCP connection with the echo server.

For an IPv4 CPE, the public routable address available in the "IP echo" measurements will be either the same address assigned to the probe, or an external-facing CPE address (in case of local NAT), or an external-facing CGNAT address. In a residential IPv6 network, a routed prefix is delegated to the CPE, often recommended to be a larger allocation than a /64, as discussed in Section 2.1. The CPE then advertises /64s from that space to devices on the home network, and RIPE Atlas probes use SLAAC to complete the remaining 64 bits. Thus, the "IP echo" measurements discover the routed LAN prefix observed by RIPE Atlas probes, and so our analyses of the temporal and spatial stability of probes' /64 prefixes also shed light on the stability of the prefix from which residential devices obtain their publicly routed IPv6 addresses.

Inferring assignment changes. Using the "IP echo" measurements, we detect assignment changes for a given probe by identifying when the reported IPv4 address (or /64 IPv6 prefix) differs from the previous one. We infer the duration of an assignment by calculating how long the assignment was continuously observed between changes. Since we restrict ourselves to observing durations only when an assignment is sandwiched between changes, we observe the exact duration (at hourly granularity) of an assignment.

Dataset characteristics. For this study we collected all available IP echo measurements from September 1, 2014 to May 31, 2020. RIPE Atlas has a well-known "geek bias" with probes sometimes being deployed in atypical manners: we account for anomalous probes and instances of spurious assignment changes using guidelines from prior work [34] (we provide details in the Appendix). In total, we found 15,982 probes that had each been observed in a single AS for longer than a month.

Of 15,982 probes, we observed at least one instance of an assignment change (in IPv4 or IPv6) in 9,448 probes. Our analyses in the rest of the paper focus upon assignment dynamics for these 9,448 probes but we briefly discuss the remaining ones here: 17% of these probes in IPv4 (22% in v6) were observed for less than 3 months; longer observation periods may have allowed us to observe assignment changes. However, 45% of these probes in IPv4 (44% in v6) did not observe an assignment change over more than a year's worth of observation; these probes' addresses may be statically assigned.

In spite of RIPE Atlas' smaller footprint compared to the CDN dataset, some residential ASes have a relatively large number of probes, which we use to observe potentially widespread patterns. Where possible, we corroborate our observations with the CDN dataset. Table 1 shows ten ASes with more than 40 dual-stack probes[3] that observed at least one address change. We expect that probes in these ASes are deployed in the home network behind the CPE and that IP Echo measurements reflect the IPv4 address (or IPv6 prefix) shared by other residential devices. We illustrate our findings from RIPE Atlas using these ASes and, where applicable, point out observations also from other ASes. The raw "IP echo" datasets are publicly available [48, 49]. We make our analysis scripts and inferences from the full dataset also publicly available [40].

[3]Each dual-stack probe yielded more than a month of IPv4 and IPv6 "IP echo" measurements.

3.2 Analyzing assignment durations

We now investigate how long CPEs retain their IPv4 and IPv6 assignments. In IPv4, the scarcity of addresses could lead ISPs to aggressively reclaim addresses through the use of short lease-times. In IPv6, on the other hand, recent recommendations are that IPv6 delegated prefixes should remain persistent [60]. We empirically examine IPv4 and IPv6 assignment durations to understand ISP practices. Our results show how long IPv4 (IPv6) addresses (prefixes) can be used to identify subscribers, and can therefore inform several applications while also raising privacy concerns.

3.2.1 Metric. Like prior work, we use the *total time fraction* metric to analyze address assignment durations [34]. Naive analysis of raw address assignment durations will overrepresent short address durations. Consider CPE_1 whose addresses typically change after a 24 hour period and CPE_2 (in the same ISP) whose addresses change typically after a 30-day period. If both CPEs have been monitored for a year, we would obtain 365 samples of 1-day durations and 12 samples of 30-day durations. The distribution of address durations from these two CPEs would overrepresent CPE_1. To avoid overrepresenting CPEs with short address durations, we use an alternative (weighted) probability mass function called the total time fraction, where instead of dividing the number of occurrences $n\ d$ of duration d by the total number of all durations in the observed population (as in a conventional probability mass function), we compute it as:

$$f_p\ d\ = n\ d \quad d\ \Sigma\ D \tag{1}$$

where D is the array of address durations from a probe (or group of probes) p, and $n\ d$ the number of times that the probe had an address duration d.[4] We then plot the associated cumulative distribution function, which we call *cumulative total time fraction*.

IPv6 assignment durations are longer than IPv4. Figure 1 shows the cumulative total time fraction for 6 ASes containing many RIPE Atlas probes. These ASes serve customers in different countries (Table 1). We split IPv4 durations into non-dual-stack and dual-stack. A probe's IPv4 duration is considered to be dual-stack if the probe has been consistently reporting IPv6 "IP echo" measurements during the same period. We observe:

> IPv6 prefixes tend to be assigned to CPEs for months by 4 of these ISPs (all but DTAG and Proximus). On the contrary, IPv4 address durations tend to be shorter, particularly for DTAG, Orange, and BT.
>
> Well-defined modes—at 1 day (DTAG), 1.5 days (Proximus), 1 week (Orange), and 2 weeks (BT)—in IPv4 non dual-stack address durations suggest that ISPs renumber addresses periodically. This result, using 6 years' worth of "IP echo" data, is consistent with observations from the 1-year "connection logs" dataset used in prior work that also noted periodic renumbering within these ISPs [34]. In total, we observe evidence of consistent periodic renumbering on 35 networks when considering non-dual-stack probes.
>
> DTAG appears to renumber IPv6 prefixes after 1-day durations as well but this pattern is not evident in the other ISPs' curves. We observe evidence of consistent periodic

[4]This is equivalent to the probability of observing a CPE with an address assigned to last d when observing a random CPE for a time frame of the same duration (d).

AS	ASN	Country	All probes	All v4 changes	Dual-stack (DS) DS probes	DS v4 changes	DS v6 changes
DTAG	3320	Germany	589	218655	402	111361 (51%)	119466
Comcast	7922	U.S.	415	4441	283	1243 (28%)	2457
Orange	3215	France	425	40085	236	4189 (10%)	746
LGI	6830	many	445	17865	141	11345 (64%)	616
Free SAS	12322	France	138	1184	90	494 (42%)	98
Kabel DE	31334	Germany	152	2525	84	1096 (43%)	173
Proximus	5432	Belgium	114	18533	64	3254 (18%)	2930
Versatel	8881	Germany	80	39110	57	30695 (78%)	31991
BT	2856	U.K.	170	15743	58	3714 (24%)	290
Netcologne	8422	Germany	43	23069	40	19223 (83%)	17087

Table 1: Overview of assignment changes observed in the RIPE Atlas "IP echo" dataset for 10 ASes with many dual-stack probes.

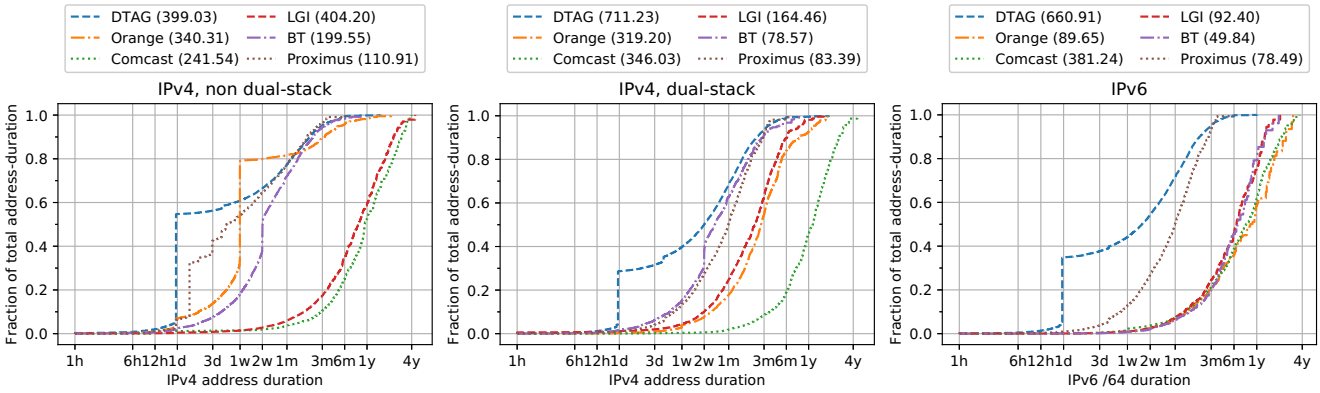


Figure 1: Cumulative total time fraction for IPv4 and IPv6 assignment durations in six large ASes. For IPv4, we further separate address durations into dual-stack (D) and non-dual-stack (ND) to highlight differences. The number in parentheses is the total assignment duration in years from all probes in the AS.

renumbering every 24 hours in IPv6 mainly in the following German ASes: DTAG, Versatel (AS8881), Netcologne (AS8422), Telefonica DE (AS6805), and M-net (AS8767). We also observe consistent period renumbering with a 12-hour period in ANTEL (AS6057) in Uruguay and with a 48-hour period in Global Village (AS18881) in Brazil.

Long IPv6 /64 durations in most ASes suggest that a /64 can be used to identify a subscriber over several months, and sometimes even years. Even if end-user devices regularly change their 64-bit host-part using temporary addresses [18, 32], the relatively static 64-bit network part permits subscriber-identification over long periods.

Probes in dual-stack networks observe longer IPv4 address durations. The two leftmost graphs in Figure 1 suggest that *IPv4* address durations are typically longer for dual-stacked hosts, when compared to non dual-stack hosts in the same network. For Orange, the difference in address durations is large: addresses assigned to dual-stack probes last significantly longer (and do not appear to change after 7-day durations). Dual-stack address durations are longer for DTAG as well, although some probes continue to have their addresses change every 24 hours even when they have dual-stack capability. We observe this general trend of longer lasting dual-stack IPv4 address durations in most networks. We also performed preliminary investigations into whether IPv4 and IPv6 assignments in dual-stack networks change simultaneously and find that the behavior varies considerably among networks; for example, in DTAG, we find the vast majority of assignment changes (90.6%) take place nearly simultaneously (we observe the changes in the same hour). We find the reverse to be true in Comcast: most changes in assignment in both IPv4 and IPv6 did not co-occur.

These analyses show that dual-stack assignment durations in both IPv4 and IPv6 can be long, often remaining stable over several months. With many networks increasingly introducing dual-stack, the stability of address durations may increase over time. These observations suggest that IP addresses can indeed be used to identify end-hosts for long periods; we discuss the applications and implications of these results further in Section 6.

Evolution over time. We next investigate how IPv4 and IPv6 assignment durations have evolved over time. For this purpose, we break down durations from each AS by year and investigate the cumulative total time fractions per year. The year-to-year trends confirm our insights from the overall dataset: IPv6 durations have consistently been longer than IPv4 durations and address durations in dual-stack networks tend to be longer than durations in non-dual-stack IPv4 networks [40]. However, we also find that assignment

durations across all categories (non-dual-stack, dual-stack, and IPv6) have shown signs of increase over the years, especially in ISPs such as DTAG and Orange which used to have short assignment durations.

Comparisons with prior work. Several prior studies have found that IPv4 addresses can change after specific periods [29, 34]. We confirm widespread periodic reassignment in 35 ISPs when considering non-dual-stack probes, although this practice seems far less common for dual-stack IPv4 probes.

Richter et al. reported upon a German ISP with 24-hour lease times and a U.S. ISP with very long lease times [45]. In general, we find 24-hour lease times to be common in German ISPs (including in dual-stack and IPv6 networks). We also find relatively long assignment durations in the U.S; probes in ISPs such as Charter, Cox, AT&T, and Time Warner have similar assignment durations when compared to those for Comcast.

On the other hand, our findings of assignment durations are longer than those from Moura et al.'s study [30]. While they found on average that IP address renewal would occur "every 61, 20, 10, and 14 hours for AT&T, British Telecom, Deutsche Telekom, and Orange, respectively", we find significantly longer durations for all these ISPs, even for non-dual-stack probes. However, given that Deutsche Telekom and other German ISPs' policy of renumbering every 24 hours has been well studied [34, 45], we suspect that the inconsistencies arise due to the Zmap-based technique's tendency to under-report session durations.

Plonka and Berger found more than 100 million IPv6 /64 prefixes to have been stable across an entire year (between March 2014 to March 2015) [36]. Our results using the RIPE Atlas dataset confirm that IPv6 /64 prefixes tend to be stable for months and years in various ASNs, although we find evidence of periodic renumbering in a handful of ISPs.

4 IPV4-IPV6 INTERPLAY

So far, we have studied temporal properties of IPv4 and IPv6 assignments using RIPE Atlas probes. In this section, we leverage a much broader, albeit less detailed, dataset captured at a major CDN. We utilize this data to corroborate earlier findings about address assignment practices, as well as to illuminate the interplay between both address types on a broad scale. In particular, we are interested in how stable associations between IPv4 and IPv6 address ranges are, and the cardinality of these relationships. Understanding stability of such associations is vital for approaches that attempt to derive knowledge about the IPv6 space by looking at their IPv4 counterparts (e.g., geolocation). Studying the cardinality of IPv4-IPv6 relationships, i.e., whether these associations are one-to-one or one-to-many, helps illuminate properties of the addressing and transitioning mechanisms used.

4.1 CDN IPv4-IPv6 assocation dataset

For this portion of our study we leverage the vantage point of a large CDN. On the one hand, the CDN dataset does not allow for detailed analyses that require pinpointing individual subscribers as the RIPE Atlas dataset does. On the other hand though, it allows us to sample the Internet much more broadly—including cellular networks, which account for a significant share of today's IPv6 adoption—and to reason about cardinalities of associations between IPv4 and IPv6 addresses.

We utilize 5 months of data from a Real-User Monitoring (RUM) system provided by the CDN. The RUM system is an optional feature used by a subset of CDN customers. The system is Javascript based, meaning responding records are sourced from Web browsers with Javascript enabled, accessing these customer pages. From these RUM transactions, we are occasionally able to extract *address associations* from particular clients, where both an IPv4 and IPv6 address is present from the same transaction. This occurs when the IP protocol used by a dual-stacked client to access the content page differs from the one used to report to the RUM server, each of which are recorded. With this, we are able to draw an instantaneous association of an IPv4 address to an IPv6 address of a single client in time. Our CDN dataset aggregates IPv4 addresses to /24 prefixes, and IPv6 addresses to /64 prefixes, and we define each association tuple as *(IPv4 /24 prefix, IPv6 /64 prefix, date)*. We gathered 5 months of address association data, between January 1, 2020 and June 1, 2020, collecting 32.7 billion associations over this period.

Pre-processing association data. For each address association, we determine the ASN of each address using BGP feeds received at the CDN. We discard any association where the ASN of the IPv4 and IPv6 address do not match. This removes both instances of multi-homed hosts, and greatly reduces the impact of potential spurious associations resulting from clients switching between networks while initiating subsequent connections to the CDN, e.g., smartphones switching between cellular and WiFi connectivity. After filtering, we are left with 31.6 billion IPv4 and IPv6 address associations during this period, observing 2.1 billion unique /64 prefixs spread across 7,775 ASNs. In light of the prevalence of IPv6 in mobile networks worldwide, we label each prefix as *mobile* for those identified as coming from cellular access networks, or *fixed* for non-cellular access. We use a similar methodology to Rula et al. [50] to identify cellular access prefixes. We find the address behavior of mobile addresses to differ greatly from that of fixed addresses, both with regard to duration and overall count. Fixed associations last 60x longer at median, and overall 65.7% of unique /64 prefixes in our dataset come from cellular access.

Limitations of CDN Data. As shown in the prior section, not all operators synchronized changes across IPv4 and IPv6 addresses. Our CDN data captures the lower bound of these changes, which are determined by a change to either address. We argue that the prefix level aggregation of our data is still representative of client address durations in many networks. The RIPE Atlas data measured assignment changes at the /64 prefix granularity by definition. Further, we show in Section 5 that in the vast majority of cases, a change in the assigned IPv4 address also results in a change in the assigned /24 prefix.

4.2 A global picture of IPv6 address durations

We next look at the duration of address associations worldwide. Differently from the RIPE Atlas data, our CDN data contains no host-level identifiers to track address changes, so we measure *association duration* as the period in which an IPv6 /64 prefix reports the same IPv4 /24 prefix. This duration is determined by the lifetime of an IPv6 /64 prefix or the appearance of another IPv4 /24 prefix.

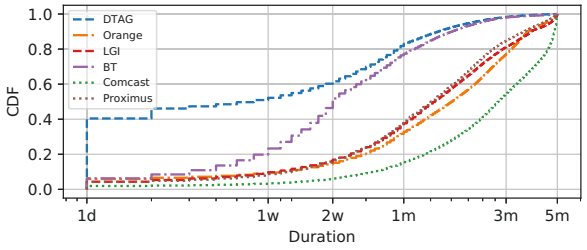


Figure 2: Address association durations for selected ISPs, observed in the CDN data. Durations closely resemble the dual-stack IPv4 address durations derived by RIPE Atlas probes.

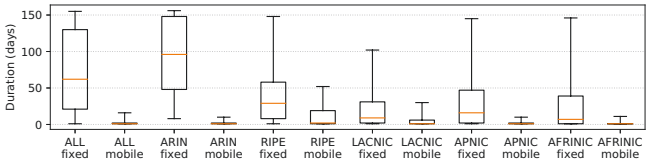


Figure 3: CDN address duration between Internet registries. The boxes represent inner quartiles, the orange line the median and the whiskers the 5th and 95th percentiles.

Comparing to the ASes from the previous section, address association durations track closely with individual host address durations. Figure 2 plots the distribution of address association duration, using the same set of operators featured in the previous section (Fig. 1). Association durations closely resemble the shorter IPv4 addressing time frames in these operators, as would be expected, since each have shorter IPv4 assignment durations than their IPv6 counterparts in general. The median association duration for DTAG and BT closely match the RIPE Atlas measurements, with durations of roughly 1 and 2 weeks, respectively. The remaining four operators show shorter durations than those derived from Atlas, with median values approximately 1-3 months less, though with relatively similar distributions.

While addressing policies can differ widely across networks, we discover that the address behavior between mobile and non-mobile addresses (i.e., fixed) are stark enough to merit analyzing separately. At a high level, we find fixed addresses to tend towards long associations, lasting around 2 months at the median. Mobile addresses, on the other hand, have a majority of associations lasting one day or less, with virtually no associations lasting longer than 30 days. Figure 3 plots the two distribution, ALL-fixed and ALL-mobile, for all networks and associations in our dataset.

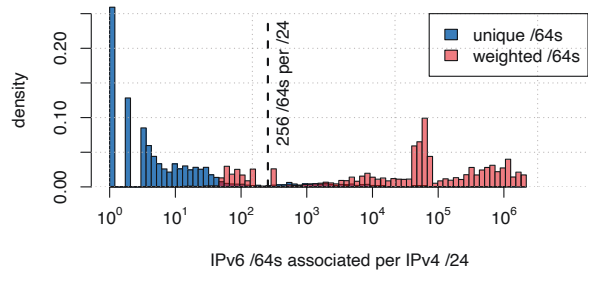
Residential IP addresses (fixed) act as stable host identifiers across both protocols over the course of months in many networks. The median association duration globally is 61 days, and 20% of associations lasted more than 143 days out of a possible 150. We believe longer measurement periods of our CDN data would yield longer durations comparable to the RIPE Atlas dataset. Conversely, the address associations observed from mobile addresses are relatively ephemeral, with 75% lasting for one day or less. The remaining 25% exhibit a long-tail lasting up to 30 days. We observe this tail behavior in all the individual mobile networks we investigate, indicating this is a property of cellular addresses and not the policy of a handful of large operators.

Geographic Considerations. We next look at the variations in association durations between geographies, grouping addresses by their delegating Internet registrar. We find rough consistency in the association behavior across geographies. Figure 3 plots the distribution of durations for each registry, split across fixed and mobile addresses. While each registrar is roughly consistent within their connection class, we observe a few regional outliers. Fixed line addresses in ARIN have very long and stable durations; the median duration is 100 days which nears our overall measurement period. Similarly, mobile address durations have inner quartiles below 5 days, with medians very near one day for ARIN, APNIC, LACNIC and AFRINIC. The main outlier, RIPE, shows a 75th percentile of 22 days. We investigated further, finding a large British mobile operator, EE Ltd., with address durations reaching up to 50 days. Since the distributions combine all addresses together, this one large operator shifts the distribution tail for the entire registrar. While our analyses from RIPE Atlas inform us that the durations of associations in fixed networks is likely bounded by the duration of the IPv4 address (since IPv6 assignments tend to last longer), we do not know if short association durations in mobile networks are caused by short IPv4 or IPv6 assignments. Since RIPE Atlas currently has limited deployment in cellular networks, we defer further exploration of this topic to future work.

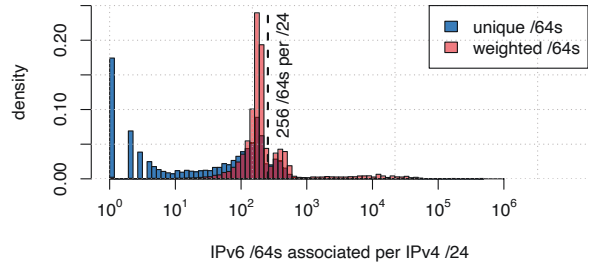
4.3 Relation between IPv4 and IPv6 assigned addresses

Next, we study the cardinality of associated IPv4 and IPv6 prefixes, by looking at the number of associated IPv6 /64 prefixes per IPv4 /24 prefix, essentially measuring the connectivity degree of each IPv4 prefix. IPv4 prefixes with high IPv6 connectivity degrees are indicative of IPv4 multiplexing through techniques such as CGNATs. We plot the distributions of unique /64 associations per /24 prefix in Figure 4, grouped by connectivity type. The figure displays the distributions of both the overall unique /24 prefixes, as well as a hit weighted distribution to account for the greater impact of highly multiplexed address blocks.

Again we find two distinct behaviors between fixed and mobile addresses. Mobile prefixes (Fig. 4a) show multiplexing to be the norm. We find the overall weighted peak at 80,000 unique /64 prefixes per IPv4 /24 prefix, and a secondary peak at just over 100,000 unique /64 prefixes. While changes to a different multiplexed IPv4 address will result in the short association durations observed in the previous section, we believe that in many networks mobile IPv6 addresses have an affinity to an IPv4 address. When looking at the inverse of the connectivity of /64 prefixes, we find that 87% of unique /64s have a connectivity of one (figure not shown). Association degree in fixed networks (Fig. 4b) shows very little evidence of multiplexing. The figure shows a weighted peak at some 150–200 unique /64s per /24 prefix, which aligns well with the typical number of active IPv4 addresses in individual /24 address blocks in residential networks [45]. This strengthens the observation that CPEs in fixed networks have stable one-to-one relationships between IPv6 and IPv4 addresses.



(a) Mobile /24 degree.



(b) Fixed /24 degree.

Figure 4: Distribution of IPv6 /64 associations with each IPv4 /24.

5 ADDRESSES IN SPACE

In this section, we study the spatial stability of public IPv4 and IPv6 assignments to subscriber networks using both the RIPE Atlas and CDN datasets. The RIPE Atlas dataset offers a unique vantage point to study pools from which CPE addresses (in IPv4) and CPE LAN prefixes (in IPv6) are allocated: since the dataset contains associations between unique probe IDs as their corresponding assignments change over time, it is possible to observe the sequence of assignments that have been assigned to a single probe. We introduce two heuristics that allow us to reason about address pools in the IPv6 space, as well as a nascent method that allows us, for some ISPs, to infer prefix lengths attributable to individual subscribers. We use the CDN dataset to corroborate and extend our findings about inferred prefix lengths.

5.1 Subsequent IPv6 assignments are typically from the same BGP prefix

Prior work has shown that subsequent assignments of IPv4 addresses to the same subscriber may come from different address blocks [34], potentially due to IPv4 address space fragmentation. Since the IPv6 address space is vast, the pool of available IPv6 addresses may exist within one contiguous address range, and so we hypothesize that subsequent IPv6 addresses assigned to the same subscriber are more likely to belong to the same block.

We quantify how often subsequent assignments to the same probe come from different BGP prefixes in IPv4 and IPv6 in Table 2 (in the Appendix). We confirm prior work's observations that subsequent IPv4 assignments often come from different /24 blocks, and even different BGP prefixes [34]. The majority of ASes in Table 2 (with the exception of Comcast, DTAG, and LGI) observe more than 40% of address changes across BGP prefixes. However, subsequent assignments are typically from the same BGP prefix in IPv6.

5.2 IPv6 Subscriber Pool Boundaries

Although IPv6 assignments to the same subscriber tend to belong to the same BGP prefix, IPv6 BGP prefixes can be massive (for example, DT announces 2003::/19). It is therefore more likely that ISPs utilize internal delineations of this address space for dynamic addressing for subscribers within a region and/or particular service tiers, and that a given CPE in the ISP would receive subsequent delegated prefixes that are spatially close to each other. If so, subsequent assignments to a CPE may occur primarily within a significantly smaller address range than the publicly announced BGP prefix.

Here, we study spatial properties of assignments to individual probes, i.e., how "far apart" /64 prefixes of one specific CPE are in the IPv6 space.

Spatial distance of successively assigned assignments. We first investigate the spatial distance between *successive* assignments to the same CPE, since it is possible that an assignment and the assignment immediately prior may share more spatial similarity. By identifying the number of bits that are likely to be common between subsequent assignments, IPv6 active probing techniques can reduce the search space among possible prefixes to find active devices, even after the delegated prefix that previously contained an active device has changed. Our metric for determining the spatial distance between successive prefixes allocated to the same CPE is the "Common Prefix Length", or CPL, between successive /64 prefixes. We define the CPL to be the number of bits from the left that are identical between successive assignments. For example, if a probe's /64 changed from 2604:3d08:4b80:aa00::/64 to 2604:3d08:4b80:aaf0::/64, then the common prefix length between these two addresses is 56. The orange bars in Figure 5 show the number of assignment changes, arranged by the common prefix lengths of successive /64s, for a selection of ISPs. The blue bars indicate the number of probes that observed at least one assignment change where successive /64s had n bits in common.

We observe that, overall, the vast majority of successive /64 prefixes from most probes share at least 40 bits in common with each other. However, the picture sharpens when we inspect individual ISPs. Since probes in DTAG (AS3320, Figure 5b) observe frequent assignment-changes even in IPv6, we have a large sample over the data collection period (2014 – 2020). We see that there are no changes where the CPL is shorter than 24. A few cases have CPLs from 24 to 40, but most of the assignment-changes have CPLs of 41 to 47 bits, and more than 100 probes contribute at least one sample with these CPLs. However, we also observe numerous assignment-changes with CPLs greater than 56. In fact, close to 100 probes contribute at least one change with a common prefix length larger or equal to 56.[5] The key takeaway here is that if subsequent assignments share 56 or more bits in common, a quick search of the neighboring 255 /64s will suffice to find a device even after an assignment-change—a useful insight for active probing approaches.

On the other hand, for Comcast (AS7922), we have considerably fewer observations and the common prefix length between consecutive assignments appears to often also be /40, as seen in Figure 5a.

[5] We speculate that these changes are not actual re-assignments coming from the ISP, but rather the result of some home CPEs that periodically scramble the available bits in the ISP-delegated prefix, a feature of many DTAG CPE devices [25], and we will further investigate this phenomenon in Section 5.3.

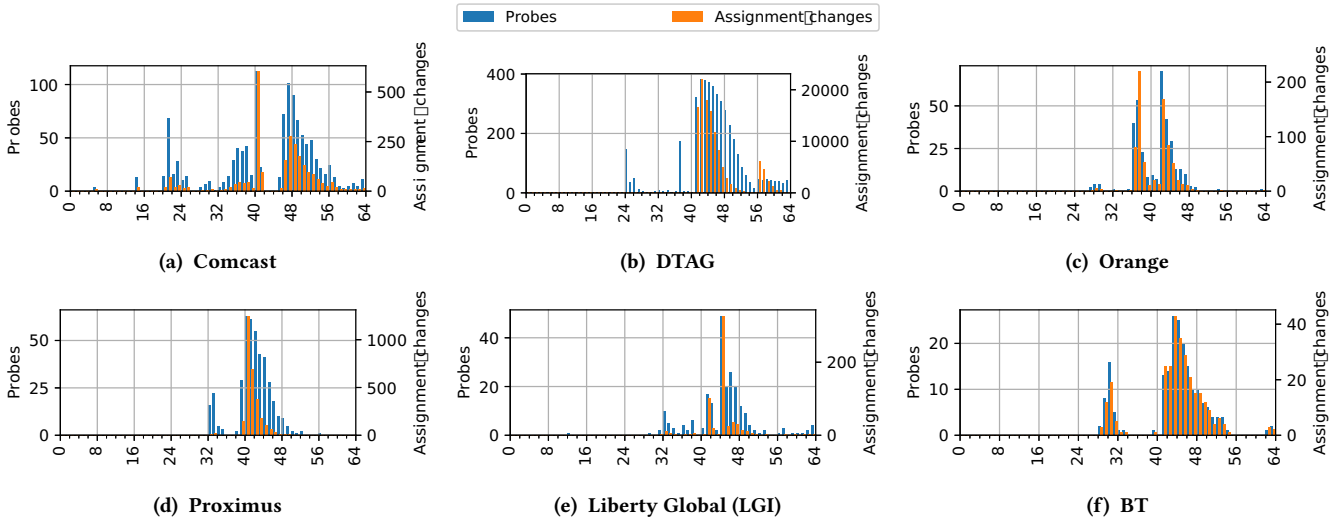


Figure 5: Common prefix lengths between subsequent IPv6 /64 prefix assignments observed by RIPE Atlas probes. The orange bars indicate the total assignment change instances where the previous and subsequent assignments shared a prefix length of *n*, i.e., subsequent addresses have *n* bits in common. The blue bars indicate the number of probes that observed at least one assignment change with *n* bits in common.

Other ASes have different behaviors: Liberty Global (AS6830) often assigns consecutive assignments that share 44 bits. In Orange, CPLs tend to cluster between bits 36 to 48. BT has two modes: one between 28 to 32 and another between bits 41 to 54. These observations show that the search space to find a device after an assignment change can vary across ISPs.

Long-term locality of assigned addresses. Figure 5 showed that subsequent assignments to a subscriber can sometimes share fewer than 40 common bits but it is unclear if such subscribers observe many distinct /40 prefixes or if assignments switch back and forth between a handful of /40 prefixes. To answer this question, we investigate the distribution of unique prefixes of various lengths observed by each RIPE Atlas probe within different networks in Figure 8 (details in the Appendix). Most probes observe less than five unique /40 prefixes over their lifetimes although they observe considerably more /48s. These results suggest that the majority of assignments to a given subscriber in IPv6 take place within the same /40.

Given that we capture longitudinal behavior, it is likely that the boundaries shown reflect addressing inside the ISP. These boundaries may delineate different address pools out of which the ISP assigns prefixes to end-users. This insight may be useful for reputation and anonymization techniques, since for many ISPs, a /40 emerges as a common size for dynamic address pools (see, e.g., DTAG for example). Such address pools might well share important commonalities, e.g., geographic location, but at the same time aggregate a sufficiently large set of end-hosts. We leave a more thorough analysis of address pools to future work but conclude this analysis with an important implication for active scanning approaches: a device with an EUI-64 address can be almost trivially located in many domestic ISPs over long time periods, by scanning prefixes within the same /40 address block. The search space is therefore reduced from the scope of the BGP announcement (in the case of DT, 2^{64-19}) down to 2^{64-40} networks. In the following section, we constrain that search space even further.

5.3 IPv6 Individual Subscriber Boundaries

Our goal in this section is to work towards uncovering the prefix length that can identify an individual subscriber. Identifying individual subscribers in IPv6 has crucial implications for privacy, address reputation, and active probing approaches. Reputation systems attempt to assign malicious activity to individual subscribers (as opposed to a group of subscribers) and anonymization approaches need to identify aggregates that must span multiple subscribers. Active probing approaches can further reduce the search space, e.g., if a subscriber is always assigned /64 prefixes with the trailing 8 bits before the /64 boundary zeroed out, and the long-term stable prefix is a /40 (recall Section 5.2), then the search space for an EUI-64 device within the subscriber's network would be constrained to other /56s (instead of /64s) within the /40.

Approach: Finding the zero bits. Our approach is to investigate /64 prefixes which have multiple trailing zero bits, i.e., zero bits in the less-specific portion before the /64 boundary. There are two scenarios which can result in this behavior. The first (and common) scenario is that an ISP may delegate a prefix that is shorter than a /64 to individual CPEs (e.g., /56s, /48s) [60] and the CPE chooses to zero out the remaining bits and announce the lowest-numbered /64 within the delegated prefix to the local subscriber network. This behavior may be CPE-dependent; since CPEs have the freedom to choose new /64s from within these assignments, they may also opt to rotate through different /64s [25]. However, based on RFCs related to prefix delegation [57] and CPE requirements [55], we expect to observe several CPEs announcing /64s with trailing zero bits immediately preceding the /64 boundary. The second scenario

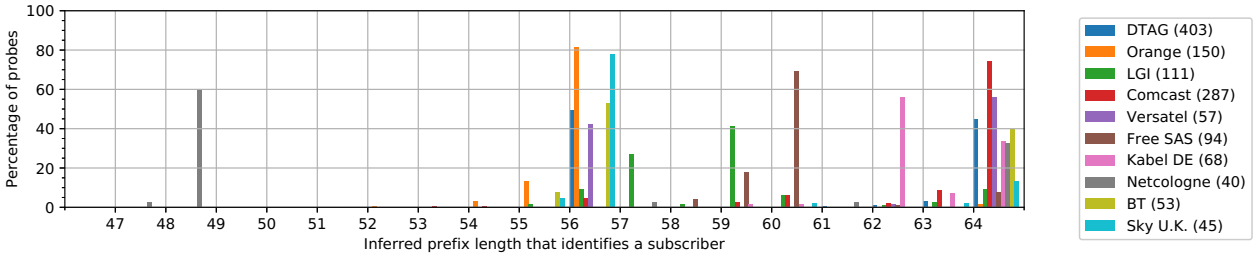


Figure 6: Inferred prefix lengths identifying a subscriber in ASes with many RIPE Atlas IPv6 probes. The numbers in parentheses next to each ISP indicate the total probes with at least one IPv6 assignment change in that ISP.

(which we expect to occur less frequently but note here for the sake of completeness) is that ISPs may be delegating /64s to CPEs but the /64s themselves may be separated by multiple bits (e.g., 16, 256), so that they always have trailing zeroes. Irrespective of the cause, the effect is that the subscriber LAN /64 prefix will have multiple trailing zeroes, and thus an even shorter prefix may be used to identify a subscriber.

Using multiple assignments assigned to the same subscriber with RIPE Atlas. Our technique for inferring subscriber-identifying prefix lengths on RIPE Atlas is to identify the number of bits upstream from bit /64 that are consistently 0 for *all* /64s we saw from a probe. We subtract this number from 64 to infer the prefix length that was likely delegated to the subscriber. Figure 9 (Appendix) shows the inferred prefix lengths for the set of all RIPE Atlas probes taken together. For about half of these probes, we indeed find less-specific prefixes with zeroed-out portions before the /64 boundary. Notably, we see a spike at the /56 boundary, which is a common prefix length that ISPs delegate to their residential subscribers [60].

Figure 6 breaks our results up for individual ISPs. Here, we can see that Orange, DTAG, as well as Sky UK show strong concentrations in /56. We were able to verify that all three ISPs indeed delegate a /56 prefix to their subscribers [22, 23, 61]. We were further able to verify Kabel DE (peak at /62): the ISP allows the CPE to request a delegation of up to a /56, but their support website states that their branded CPE devices request a /62 delegation [11]. Netcologne has several probes with /48 prefix lengths (see grey bar towards the left); we were able to verify that Netcologne indeed delegates entire /48 prefixes to individual subscribers [33]. The fact that subscribers can be delegated such short prefixes has vital implications for anonymization approaches: if an approach chooses a /48 boundary to identify a set of users (as, e.g., Google Analytics does [21]), this set would consist of a single subscriber in the case of Netcologne! Plonka and Berger had found that a Japanese ISP likely delegates /48 prefixes to subscribers [37]; our results show that this practice occurs in other ISPs too.

These findings suggest that our method can infer the prefix length identifying individual subscribers in many instances. However, this method only works for CPE devices that zero out available bits. If CPE devices scramble the available bits (as some CPEs in DTAG do [25]) or use non-zero constant identifiers, our approach may overestimate the prefix length. This is exemplified by the second spike for DTAG at the /64 boundary. In such cases, we cannot determine if the non-zero bits after the /56 are due to ISP assignment policy or CPE prefix scrambling policy. Contrarily, if we observe relatively few assignment changes for a probe, the set of /64s from that probe may share more zero bits than the delegated prefix by random chance. However, the likelihood of inferring shorter prefixes is low; e.g., even if we observe only two /64 prefixes, the probability that both addresses have 0s in their last 8 bits is very small.

Using multiple addresses observed at the large CDN. We utilize our CDN dataset to measure the validity of this technique at a global scale. Looking at all collected /64 IPv6 prefixes from the CDN dataset, we calculate the prevalence of trailing zeros, and its ability to act as a general technique for detecting delegated prefix lengths. We observe a high frequency of trailing zeros across the IPv6 address space in fixed addresses, with certain operators showing highly consistent behavior in this addressing pattern. For example, Orange in France has IPv6 /64 prefixes with the last 8 bits as zeros in 99.7% of /64 prefixes. Mobile /64 prefixes show no evidence of consistent trailing zeroes, suggesting that mobile subscribers are typically delegated /64 prefixes.

For fixed addresses, we measured the extent of this effect by looking at the fraction of addresses with trailing zeros, and classifying addresses by their longest streak of zeros across consecutive nibble boundaries. For example, an address with the last 8 bits as zeros would match the /56 boundary, whereas an address with the last 16 bits as zeros would match the /48 boundary. We plot the frequency of these trailing zeros, grouped by their longest prefix boundaries, for all observed /64 prefixes in Figure 7. The bars in the figure correspond to the fraction of addresses in each registry which end in trailing zeros beyond each prefix boundary, and correspond to the fraction of detectable delegated prefixes at each length. In ARIN, for instance, we find that 59% of all /64s show either 4, 8, 12, or 16 trailing zeros. At individual prefix levels, we find 30% of observed /64s have zeros only in the last four bits, allowing us to infer a /60 delegated prefix length while 27% of prefixes have zeros only in the last 8 bits, signifying /56 prefix delegations.

We find this technique to be widely applicable across the globe: 43.2% of all fixed /64 prefixes contain trailing zeros which allow an inferable prefix delegation. We find regional patterns exist across regions, with all but LACNIC containing 54.5-83.1% of addresses which have inferable prefixes. Certain geographies such as RIPE and AFRINIC have very specific patterns, with over 60% of all /64 prefixes with zeros in trailing 8 bits, a /56 delegation.

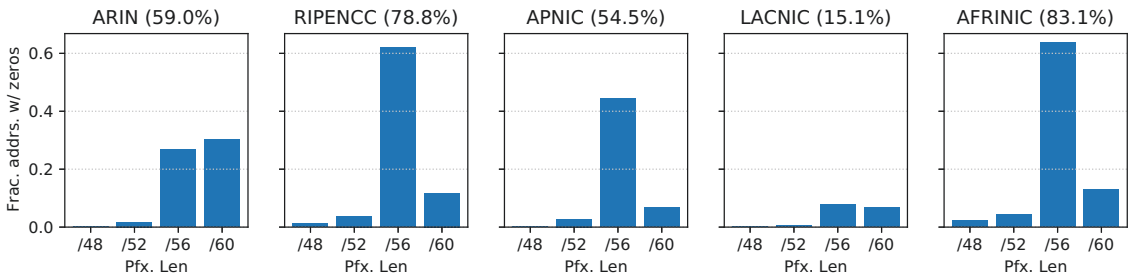


Figure 7: Observed frequency of trailing zeros used to infer delegated prefix lengths. Bars signify the fraction of observed /64 prefixes which contain zeros to the right of the inferred prefix length. Addresses are classified by their longest string of zeros. Percentages in title refer to percent of addresses we were able to infer prefix delegation length with this method.

6 IMPLICATIONS AND APPLICATIONS

Our observations of assignment practices offer insights that can benefit a variety of applications. Our findings about long-lasting dual-stack IPv4 and IPv6 assignment durations show that it may be possible to use addresses as host identifiers for long periods. Determining IPv6 address pool boundaries and inferring IPv6 prefixes delegated to a subscriber helps applications reason about what IPv6 prefixes of various lengths represent.

Active scanning and target generation. Emerging active scanning approaches for the IPv6 space rely on heuristics and structure embedded in IP addresses to reduce the search space to make active probing feasible [3, 17, 31]. Our findings on key aspects of address assignment policies for subscribers can provide concrete data on address ranges in which legitimate targets may reside. First, in Section 5.2, we present data on the typical size of the pool from which a network draws its subscriber allocations for domestic networks; these are often /40, but vary per network. Second, in Section 5.3 we show typical sizes of the address block allocated to CPEs for subsequent use in various subscriber networks; our initial findings suggest /56s and /60s (among others) are commonplace. IPv6 host addressing continues to evolve and makes individual hosts difficult to track across networks. Regardless, EUI-64 IIDs are still commonplace, and in some cases (as with RIPE Atlas probes), are intended to be stable to facilitate their use as reliable measurement targets. Although individual networks choose their own addressing plans, knowing a network's active usage is valuable for constraining the search space of IPv6 scanning tools and IPv6 hitlists, reducing the number of redundant probes.

Host reputation systems. Host reputation systems rely on the notion of an IP address to attribute malicious activity to individual hosts [5, 26, 27, 41, 42]. Being able to determine IPv6 prefix lengths that identify individual subscribers is vital to avoid collateral damage, i.e., erroneously attributing malicious activity to a larger set of users, and to avoid evasion, i.e., attributing malicious activity to a too-specific prefix, which the host in question can easily change. Information on the typical address block size allocated to CPEs, as we presented in Section 5.3, is critical for this application. Also critical is the meaningful duration before a bad actor migrates to a new address. Section 3.2 shows how long assignments typically stay active. We observe variation per-ISP but also identify common patterns that hold across ISPs: many IPv6 assignments are long-lived, though DTAG in particular demonstrates assignment durations of 24 hours. Additionally, we observe a distinct difference in IPv4 behaviour between single-stack and dual-stack networks. In Section 4.2 we corroborate our RIPE Atlas findings with observations from a global CDN. Since dual-stack networks are becoming common, a bad actor is likely to target services over IPv4 and IPv6. To that end, in Section 4.3 we indicate that in fixed-line networks there is a strong association between IPv4 assignments and IPv6 assignments: we observe many IPv4 /24s with 150 – 200 distinct IPv6 /64s, in line with a typical IPv4 NAT configuration.

User Privacy and Anonymization techniques. Our results on the long duration of delegated IPv6 prefixes to subscribers show that the current standards for privacy-enhancing addressing in IPv6 (such as RFC 4941 [32]) are not sufficient to protect users' privacy. Furthermore, they highlight—as suggested also in [37]—that *simple* anonymization by truncation [21] is fallacious, since it does not account for the diversity in address assignment practices we observe (such as the delegation of /48 prefixes to individual subscribers). Anonymization techniques for sharing data containing IPv6 addresses must rely on knowledge of prefix boundaries that identify individual subscribers, or subscriber pools, in order to aggregate potentially sensitive information so that individual subscribers cannot be identified [37]. While in the IPv4 space, aggregating addresses to a /24 prefix is a common technique [37], boundaries in the IPv6 space depend on individual ISPs and their assignment practices. Our findings bring us one step closer to public, data-driven metrics that may allow a per-network approach to obfuscating IPv6 datasets, with the added benefit of facilitating data sharing for research.

7 CONCLUSION

In this work we used complementary datasets from RIPE Atlas and a large CDN to investigate temporal and spatial dynamics of IPv4 and IPv6 address assignments. We found that IPv6 assignments typically last longer than IPv4 assignments and can persist for months in several large residential ISPs. We studied spatial aspects of IPv6 addresses in detail, identifying subscriber pool boundaries, as well as individual subscriber boundaries. We believe that our results can serve as viable input for active probing approaches in the IPv6 Internet and host reputation systems, and provide empirical data for discussion and measures to preserve privacy when assigning IPv6 prefixes to individual subscribers or when anonymizing datasets.

ACKNOWLEDGMENTS

We thank the anonymous reviewers and shepherd for their helpful feedback. This research was supported by NSF CNS-1901517, NSF CNS-1705024, and by the Open Technology Fund.

REFERENCES

[1] [n. d.]. Routeviews Prefix to AS mappings Dataset for IPv4 and IPv6. https://www.caida.org/data/routing/routeviews-prefix2as.xml. ([n. d.]).
[2] Oded Argon, Anat Bremler-Barr, Osnat Mokryn, Dvir Schirman, Yuval Shavitt, and Udi Weinsberg. 2010. On the Dynamics of IP Address Allocation and Availability of End-Hosts. *CoRR* abs/1011.2324 (2010). arXiv:1011.2324 https://arxiv.org/abs/1011.2324
[3] Robert Beverly, Ramakrishnan Durairajan, David Plonka, and Justin P. Rohrer. 2018. In the IP of the Beholder: Strategies for Active IPv6 Topology Discovery. In *ACM Internet Measurement Conference (IMC)*.
[4] Kevin Borgolte, Shuang Hao, Tobias Fiebig, and Giovanni Vigna. 2018. Enumerating Active IPv6 Hosts for Large-scale Security Scans via DNSSEC-signed Reverse Zones. In *Proceedings of the 39th IEEE Symposium on Security & Privacy (S&P)*.
[5] Qiang Cao, Xiaowei Yang, Jieqi Yu, and year = 2014 Christopher Palow, booktitle = CCS. [n. d.]. Uncovering Large Groups of Active Malicious Accounts in Online Social Networks.
[6] Martin Casado and Michael J. Freedman. 2007. Peering Through the Shroud: The Effect of Edge Opacity on IP-Based Client Identification. In *Symposium on Networked Systems Design and Implementation (NSDI)*.
[7] Jacky C. Chu, Kevin S. Labonte, and Brian N. Levine. 2002. Availability and locality measurements of peer-to-peer file systems. In *ITCom: Scalability and Traffic Control in IP Networks*.
[8] Jakub Czyz, Mark Allman, Jing Zhang, Scott Iekel-Johnson, Eric Osterweil, and Michael Bailey. 2014. Measuring IPv6 Adoption. In *ACM SIGCOMM*.
[9] Jakub Czyz, Michael Kallitsis, Manaf Gharaibeh, Christos Papadopoulos, Michael Bailey, and Manish Karir. 2014. Taming the 800 Pound Gorilla: The Rise and Decline of NTP DDoS Attacks. In *ACM Internet Measurement Conference (IMC)*. https://doi.org/10.1145/2663716.2663717
[10] David Dagon, Niels Provos, Christopher P Lee, and Wenke Lee. 2008. Corrupted DNS Resolution Paths: The Rise of a Malicious Resolution Authority. In *Network and Distributed System Security Symposium (NDSS)*.
[11] Vodafone Kabel Deutschland. [n. d.]. Haeufige Fragen (in German). ([n. d.]). https://kabel.vodafone.de/hilfe_und_service/faq
[12] Amogh Dhamdhere, Matthew Luckie, Bradley Huffaker, kc claffy, Ahmed Elmokashfi, and Emile Aben. 2012. Measuring the Deployment of IPv6: Topology, Routing and Performance. In *ACM Internet Measurement Conference (IMC)*.
[13] Ralph Droms. 1997. *Dynamic Host Configuration Protocol*. RFC 2131. https://tools.ietf.org/html/rfc2131
[14] Holly Esquivel, Aditya Akella, and Tatsuya Mori. 2010. On the Effectiveness of IP Reputation for Spam Filtering. In *2010 Second International Conference on COMmunication Systems and NETworks (COMSNETS 2010)*.
[15] Tobias Fiebig, Kevin Borgolte, Shuang Hao, Christopher Kruegel, and Giovanni Vigna. 2017. Something from Nothing (There): Collecting Global IPv6 Datasets from DNS. In *Passive and Active Network Measurement Conference (PAM)*.
[16] Pawel Foremski, David Plonka, and Arthur W. Berger. 2016. Entropy/IP: Uncovering Structure in IPv6 Addresses. In *ACM Internet Measurement Conference (IMC)*.
[17] Oliver Gasser, Quirin Scheitle, Pawel Foremski, Qasim Lone, Maciej Korczynski, Stephen D. Strowes, Luuk Hendriks, and Georg Carle. 2018. Clusters in the Expanse: Understanding and Unbiasing IPv6 Hitlists. In *ACM Internet Measurement Conference (IMC)*.
[18] F. Gont. 2014. *A Method for Generating Semantically Opaque Interface Identifiers with IPv6 Stateless Address Autoconfiguration (SLAAC)*. RFC 7217. https://tools.ietf.org/html/rfc7217
[19] F. Gont and T. Chown. 2016. *Network Reconnaissance in IPv6 Networks*. RFC 7707. https://tools.ietf.org/html/rfc7707
[20] F. Gont, A. Cooper, D. Thaler, and W. Liu. 2017. *Recommendation on Stable IPv6 Interface Identifiers*. RFC 8064. https://tools.ietf.org/html/rfc8064
[21] Google. [n. d.]. IP Anonymization (or IP masking) in Analytics. https://support.google.com/analytics/answer/2763052. ([n. d.]).
[22] C. Jacquenet and M. Sall. 2009. France Telecom's IPv6 Strategy. (Nov. 2009). https://meeting.afrinic.net/afrinic-11/slides/day1/IPv6_Strategy_Modou_SALL.pdf
[23] Reiko Kaps. 2012. Details zu IPv6 ueber Telekom-DSL (in German). (May 2012). https://www.heise.de/newsticker/meldung/Details-zu-IPv6-ueber-Telekom-DSL-1762367.html
[24] Maria Konte, Roberto Perdisci, and Nick Feamster. 2015. ASwatch: An AS Reputation System to Expose Bulletproof Hosting ASes. *SIGCOMM Computer Communication Review* 45, 4 (Aug. 2015), 625–638. https://doi.org/10.1145/2829988.2787494
[25] Stefan Krempl. 2011. Heise online: Deutsche Telekom stellt Datenschutztechnik für IPv6 vor (in German). (2011). https://www.heise.de/newsticker/meldung/Deutsche-Telekom-stellt-Datenschutztechnik-fuer-IPv6-vor-1383772.html
[26] Frank Li and David Freeman. 2020. Towards A User-Level Understanding of IPv6 Behavior. In *ACM Internet Measurement Conference (IMC)*.
[27] Vector Guo Li, Matthew Dunn, Paul Pearce, Damon McCoy, Geoffrey M. Voelker, and Stefan Savage. 2016. Reading the Tea leaves: A Comparative Analysis of Threat Intelligence. In *USENIX Security Symposium*.
[28] I. Livadariu, K. Benson, A. Elmokashfi, A. Dainotti, and A. Dhamdhere. 2018. Inferring Carrier-Grade NAT Deployment in the Wild. In *IEEE Conference on Computer Communications (INFOCOM)*.
[29] Gregor Maier, Anja Feldmann, Vern Paxson, and Mark Allman. 2009. On Dominant Characteristics of Residential Broadband Internet Traffic. In *ACM Internet Measurement Conference (IMC)*.
[30] Giovane C. M. Moura, Carlos Ganán, Qasim Lone, Payam Poursaied, Hadi Asghari, and Michel van Eeten. 2015. How Dynamic is the ISPs Address Space? Towards Internet-Wide DHCP Churn Estimation. In *IFIP Networking Conference (IFIP Networking)*. IEEE.
[31] Austin Murdock, Frank Li, Paul Bramsen, Zakir Durumeric, and Vern Paxson. 2017. Target Generation for Internet-wide IPv6 Scanning. In *ACM Internet Measurement Conference (IMC)*.
[32] Thomas Narten, Richard Draves, and Suresh Krishnan. 2007. *Privacy Extensions for Stateless Address Autoconfiguration in IPv6*. RFC 4941. https://tools.ietf.org/html/rfc4941
[33] NetCologne. [n. d.]. IPv6 bei NetCologne (in German). ([n. d.]). https://www.netcologne.de/privatkunden/ipv6
[34] Ramakrishna Padmanabhan, Amogh Dhamdhere, Emile Aben, kc claffy, and Neil Spring. 2016. Reasons Dynamic Addresses Change. In *ACM Internet Measurement Conference (IMC)*.
[35] Ramakrishna Padmanabhan, Aaron Schulman, Dave Levin, and Neil Spring. 2019. Residential Links Under the Weather. In *ACM SIGCOMM*.
[36] David Plonka and Arthur W. Berger. 2015. Temporal and Spatial Classification of Active IPv6 Addresses. In *ACM Internet Measurement Conference (IMC)*.
[37] David Plonka and Arthur W. Berger. 2017. kIP: a Measured Approach to IPv6 Address Anonymization. *CoRR* abs/1707.03900 (2017). https://arxiv.org/abs/1707.03900
[38] The Spamhaus Project. 2011. Spamhaus IPv6 Blocklists Strategy Statement. (June 2011). https://www.spamhaus.org/organization/statement/012/spamhaus-ipv6-blocklists-strategy-statement
[39] Moheeb Abu Rajab, Jay Zarfoss, Fabian Monrose, and Andreas Terzis. 2007. My Botnet is Bigger than Yours (Maybe, Better than Yours): why size estimates remain challenging. In *Proceedings of the 1st USENIX Workshop on Hot Topics in Understanding Botnets, Cambridge, USA*.
[40] Ramakrishna Padmanabhan and John P. Rula and Philipp Richter and Stephen D. Strowes and Alberto Dainotti. 2020. Paper website. (2020). https://www.caida.org/publications/papers/2020/dynamips/supplemental/
[41] Sivaramakrishnan Ramanathan, Anushah Hossain, Jelena Mirkovic, Minlan Yu, and Sadia Afroz. 2020. Quantifying the Impact of Blacklisting in the Age of Address Reuse. In *ACM Internet Measurement Conference (IMC)*.
[42] Sivaramakrishnan Ramanathan, Jelena Mirkovic, and Minlan Yu. 2020. BLAG: Improving the Accuracy of Blacklists. In *Network and Distributed System Security Symposium (NDSS)*.
[43] Y. Rekhter, B. Moskowitz, D. Karrenberg, G. J. de Groot, and E. Lear. 1996. *Address Allocation for Private Internets*. RFC 1918. https://tools.ietf.org/html/rfc1918
[44] Philipp Richter, Mark Allman, Randy Bush, and Vern Paxson. 2015. A Primer on IPv4 Scarcity. *ACM Computer Communication Review* 45, 2 (April 2015). Editorial Contribution.
[45] Philipp Richter, Georgios Smaragdakis, David Plonka, and Arthur Berger. 2016. Beyond Counting: New Perspectives on the Active IPv4 Address Space. In *ACM Internet Measurement Conference (IMC)*.
[46] Philipp Richter, Florian Wohlfart, Narseo Vallina-Rodriguez, Mark Allman, Randy Bush, Anja Feldmann, Christian Kreibich, Nicholas Weaver, and Vern Paxson. 2016. A Multi-Perspective Analysis of Carrier-Grade NAT Deployment. In *ACM Internet Measurement Conference (IMC)*. https://doi.org/10.1145/2987443.2987474
[47] Carl Rigney, Steve Willens, Allan C. Rubens, and William Allen Simpson. 2000. *Remote Authentication Dial In User Service (RADIUS)*. RFC 2865. https://tools.ietf.org/html/rfc2865
[48] RIPE NCC. 2014 – 2020. RIPE Atlas IP echo measurements in IPv4. https://atlas.ripe.net/measurements/12027/. (2014 – 2020).
[49] RIPE NCC. 2014 – 2020. RIPE Atlas IP echo measurements in IPv6. https://atlas.ripe.net/measurements/13027/. (2014 – 2020).
[50] John P Rula, Fabián E Bustamante, and Moritz Steiner. 2017. Cell Spotting: Studying the Role of Cellular Networks in the Internet. In *ACM Internet Measurement Conference (IMC)*.
[51] Erik Rye and Robert Beverly. 2020. Discovering the IPv6 Network Periphery. In *Passive and Active Network Measurement Conference (PAM)*.
[52] Quirin Scheitle, Oliver Gasser, Minoo Rouhi, and Georg Carle. 2017. Large-Scale Classification of IPv6-IPv4 Siblings with Variable Clock Skew. In *Network Traffic Measurement and Analysis Conference (TMA)*.

[53] Kyle Schomp, Tom Callahan, Michael Rabinovich, and Mark Allman. 2013. On Measuring the Client-Side DNS Infrastructure. In *ACM Internet Measurement Conference (IMC)*.
[54] Aaron Schulman and Neil Spring. 2011. Pingin' in the Rain. In *ACM Internet Measurement Conference (IMC)*.
[55] H. Singh, W. Beebee, C. Donley, and B. Stark. 2013. *Basic Requirements for IPv6 Customer Edge Routers.* RFC 7084. https://tools.ietf.org/html/rfc7084
[56] S. Thomson, T. Narten, and T. Jinmei. 2007. *IPv6 Stateless Address Autoconfiguration.* RFC 4862. https://tools.ietf.org/html/rfc4862
[57] O. Troan and R. Droms. 2003. *IPv6 Prefix Options for Dynamic Host Configuration Protocol (DHCP) version 6.* RFC 3633. https://tools.ietf.org/html/rfc3633
[58] J. Ullrich, P. Kieseberg, K. Krombholz, and E. Weippl. 2015. On Reconnaissance with IPv6: A Pattern-Based Scanning Approach. In *2015 10th International Conference on Availability, Reliability and Security.*
[59] Peter van Dijk. 2012. Finding v6 hosts by efficiently mapping ip6.arpa. https://web.archive.org/web/20170603234058/http://7bits.nl/blog/posts/finding-v6-hosts-by-efficiently-mapping-ip6-arpa. (March 2012).
[60] Jan Žorž, Sander Steffann, Primož Dražumerič, Mark Townsley, Andrew Alston, Gert Doering, Jordi Palet, Jen Linkova, Luis Balbinot, Kevin Meynell, and Lee Howard. 2017. Best Current Operational Practice for Operators: IPv6 prefix assignment for end-users - persistent vs non-persistent, and what size to choose. (Oct. 2017). https://www.ripe.net/publications/docs/ripe-690
[61] Martin Wasley. [n. d.]. Setup for Sky UK ISP. ([n. d.]). https://docs.opnsense.org/manual/how-tos/SkyUK.html
[62] J. Weil, V. Kuarsingh, C. Donley, C. Liljenstolpe, and M. Azinger. 2012. *IANA-Reserved IPv4 Prefix for Shared Address Space.* RFC 6598. https://tools.ietf.org/html/rfc6598
[63] Yinglian Xie, Fang Yu, Kannan Achan, Eliot Gillum, Moises Goldszmidt, and Ted Wobber. 2007. How dynamic are IP addresses?. In *ACM SIGCOMM.*
[64] Sebastian Zander, Lachlan L. H. Andrew, Grenville J. Armitage, Geoff Huston, and George Michaelson. 2012. Mitigating Sampling Error when Measuring Internet Client IPv6 Capabilities. In *ACM Internet Measurement Conference (IMC).*

A APPENDIX

A.1 Sanitizing the RIPE Atlas IP echo dataset

We collected all available IP echo measurements from September 1 2014 to May 31 2020 for this study. During this time, we observed at least one hourly measurement in IPv4 or IPv6 from 25,504 probes. For each assignment observed by these probes, we used the Routeviews pfx2as dataset [1] to obtain routed BGP prefixes.

We used heuristics to detect and filter probes that appear to be deployed in atypical scenarios. Such probes can lead us to infer false assignment changes: for example, a probe in a multihomed network (i.e., with IP addresses from at least 2 ISPs) may make one IP echo measurement $IP1$ (from the first ISP), the next measurement using $IP2$ (from the second ISP), and the third measurement with $IP1$ again. In such scenarios, we may incorrectly infer that the assigned IP address changed from $IP1$ to $IP2$ and back again to $IP1$ *even though* $IP1$ *and* $IP2$ *have been continuously and simultaneously assigned to the subscriber*. We filtered anomalous probes using guidelines from prior work [34] (Section 3.2) and some other heuristics we developed empirically, listed below:

Short-duration probes: Many probes only yielded measurements for short periods; we focused upon the 18,525 probes that yielded measurements for at least a month.
Multihomed probes: We filtered probes that are *multihomed*, since assignment changes on such probes are ambiguous. Probes in multihomed networks can choose any of the available addresses in their "IP echo" measurements; consequently, a change in a subsequent measurement is ambiguous. We identified multihomed probes by looking for probes that reported measurements from alternating addresses and/or Autonomous Systems. We found 5,715 probes to be multihomed and filtered them.
Bad tag probes: Since our focus is upon residential networks, we filtered probes that are obviously not in residences. Atlas users can provide "tags" associated with a probe; we filtered probes that had at least one of the following tags: "multihomed", "datacentre", "core", or "system-anchor".
Atypical NATs: Probes report their publicly visible IP address in the *X-Client-IP* field of the IP echo dataset (Section 3.1). Probes also report their currently assigned address within the home network in the *src_addr* field. Our expectation in a typical residential setting is that the probe is deployed behind a NAT in IPv4, so that it has an RFC 1918 address [43], and that the RFC 1918 address will be reported in the *src_addr* field. We therefore filtered probes that reported a publicly visible IPv4 address in their *src_addr* field in the IPv4 Echo dataset. Conversely, in IPv6, we do *not* expect the probe to be behind a NAT. Our expectation in IPv6 is that the *X-Client-IP* and *src_addr* will be identical and will reflect the publicly visible IPv6 address of the probe. Consequently, we filtered probes that reported entries where the *X-Client-IP* and *src_addr* were not equal in the IPv6 Echo dataset.

For the probes filtered using the above criteria, we do not consider *any* IPv4 or IPv6 assignments. However, for some probes, we only omit a few assignment changes that we identify as likely spurious and consider other assignment changes. Spurious assignment changes can occur when a probe is moved from one physical location to another, or when a probe's owner switches their ISP. For example, many probes had their first IPv4 address set to 193.0.0.78; this address belongs to the RIPE NCC and was used for testing probes before the probes were distributed to volunteers. We therefore filtered all IP echo entries where the IPv4 address was 193.0.0.78. Further, while investigating multihomed behavior, we found that 2,517 probes did not alternate between ASes (our heuristic for determining multihoming), but instead, switched over entirely to a different AS. These probes' assignment sequences are consistent with the scenario where the probe's owner changed ISP. We treat these 2,517 probes as multiple "virtual probes" (one per AS) and obtain 15,982 such probes that had each been observed in a single AS for longer than a month. We treat these virtual probes and other probes identically, so for simplicity we make no distinction between them in the rest of the paper.

A.2 Subsequent IPv6 assignments are typically from the same BGP prefix

Table 2 shows how often subsequent assignments to the same RIPE Atlas probe come from different BGP prefixes in IPv4 and IPv6. We confirm prior work's observations that subsequent IPv4 assignments often come from different /24 blocks, and even different BGP prefixes [34]. The majority of ASes in Table 2 (with the exception of Comcast, DTAG, and LGI) observe more than 40% of address changes across BGP prefixes. However, subsequent assignments are typically from the same BGP prefix in IPv6.

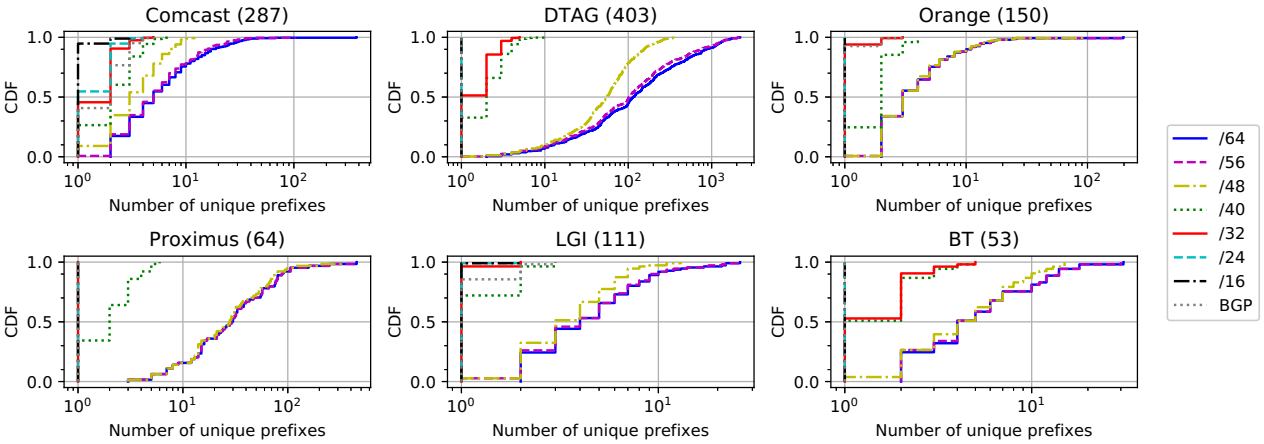


Figure 8: Cumulative distribution of unique prefixes of various lengths observed by RIPE Atlas probes. Most probes observe almost as many unique /56 prefixes as unique /64 prefixes and slightly fewer unique /48 prefixes. However, 90% of probes observe addresses from 3 or fewer /40 prefixes.

AS	Diff /24	Diff BGP (v4)	Diff BGP (v6)
DTAG	94%	27%	0%
Comcast	49%	43%	10%
Orange	99%	60%	2%
LGI	59%	14%	2%
Free SAS	100%	72%	42%
Kabel DE	84%	60%	5%
Proximus	88%	56%	0%
Versatel	93%	59%	1%
BT	94%	45%	0%
Netcologne	99%	61%	7%

Table 2: Percentage of changes in assignments across BGP prefixes. The Diff /24 column shows IPv4 address changes where the previous and next address belonged to different /24 blocks. The Diff BGP columns show the percentage of changes where the previous and the next assignment belonged to different BGP prefixes, in IPv4 and IPv6.

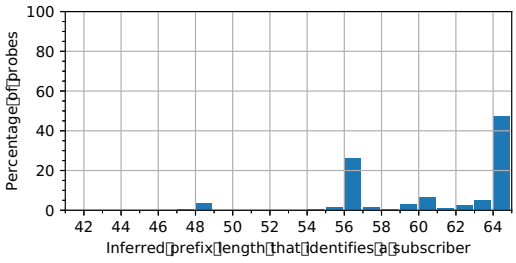


Figure 9: Inferred prefix lengths identifying a subscriber for the set of all RIPE Atlas probes with at least one IPv6 assignment change (there were 3025 such probes).

A.3 Inferred prefix lengths identifying a subscriber for the set of all RIPE Atlas probes

Figure 9 shows the inferred prefix lengths for the set of all RIPE Atlas probes taken together. For about half of these probes, we indeed find less-specific prefixes with zeroed-out portions before the /64 boundary. Notably, we see a spike at the /56 boundary, which is a common prefix length that ISPs delegate to their residential subscribers [60].

A.4 How many unique IPv6 prefixes does each probe observe?

Figure 5 showed that subsequent assignments to a subscriber can sometimes share fewer than 40 common bits but it is unclear if such subscribers observe many distinct /40 prefixes or if assignments switch back and forth between a handful of /40 prefixes. To answer this question, we investigate the distribution of unique IPv6 prefixes of various lengths observed by each RIPE Atlas probe within different networks in Figure 8. Since the number of unique prefixes may also be a function of the number of assignments observed by a probe (a probe which observes 100 changes in its assigned network may observe more prefixes than a probe with just two such changes), we also show the distribution of the number of unique /64 prefixes observed by each probe.

The /64 curve represents per-probe the distribution of assignment changes observed. In our dataset, we found that a probe is very rarely assigned the same /64 prefix again (such observations of repeated /64 prefixes happened on probes in DTAG, or Versatel, which observed hundreds of changes); thus, the distribution of the unique /64 blocks observed by probes aligns closely with the distribution of the total number of assignment changes. We observe that 35% of probes in all networks observe only a single new assignment during this study (and therefore see only two unique /64s) in Figure 8, whereas more than 50% of probes in DT observe over 100

unique /64 prefixes in Figure 8. We further observe that the distribution of /56s and /48s is similar to that of /64s, showing that (i) most new assignments are from outside these boundaries (ii) probes are *not* observing addresses from repeated /56s and /48s (if a probe was observing these prefixes repeatedly over time, the number of unique prefixes would be significantly smaller than the number of /64s for that probe). However, we observe a significant difference in the distribution of the unique /40 (and shorter) prefixes: there are far fewer of these prefixes per probe.

These results suggest that the majority of assignments in IPv6 take place within the same /40, but typically not within the same /48.