CNS Core: Medium: Collaborative Research: Strategies for Large-Scale IPv6 Active Mapping (SLAM)
Project Description

1 Introduction

Network maps – including graphs of routers, links, locations, and other meta-data – are as fundamental and crucial to network operations, engineering, and research as physical road maps are to driving. It is no surprise, then, that the networking community has expended significant effort over the past several decades developing tools and techniques, both active and passive, to gather Internet topologies (see e.g. [35, 66, 38] and references therein for an overview of this body of work). Today, several organizations, including our own, collect, infer, and annotate the IPv4 Internet topology – producing network maps that have been instrumental in supporting and advancing science, policy, research, and engineering. Internet topology data has been used in everything from censorship inference to geolocation to content distribution to security.

Less attention has been paid to the IPv6 topology. Part of the lack of focus on IPv6 has historically been due to low IPv6 adoption and use rates. However, fundamental differences from IPv4 hinder *comprehensive and representative* IPv6 topology measurement:

- **A massive address space** (2^{128}) that is sparsely populated ($\sim 2^{49}$ addresses currently advertised [36]) and cannot be exhaustively scanned. Researchers are only now experimenting with and finding viable methods for selecting measurement targets in IPv6 [29, 48, 32].
- **Mandated Rate limiting** in ICMPv6 [20] imposes a measurement catch-22: probing faster to sample more of the space is apt to induce more rate-limiting, and is thus self-defeating.
- **Address agility** that accommodates regularly shifting customers across IPv6 prefixes, and widespread use of temporary private addresses [49, 52] which makes many targets ephemeral.

As a result, today's IPv6 topology mapping systems essentially apply existing IPv4 discovery tools and techniques, e.g. continual traceroute-based probing to the base (i.e. `::1`) address of every globally advertised IPv6 BGP prefix [18, 55]. This sparse sampling of the large IPv6 address space, where prefixes are often subnetted by regions or customers, yields topology data whose completeness and quality cannot be quantified.

Our central motivating observation is that while some IPv4 measurement techniques are directly applicable to characterizing and understanding the IPv6 Internet, *many scientific, engineering, and operational planning questions require the design of new measurement strategies, techniques, and tools that explicitly consider the unique properties of the IPv6 protocol, implementations, and operational deployment [15].*

We propose to rigorously investigate, develop, and evaluate Strategies for Large-Scale IPv6 Active Mapping (SLAM), with three inter-related and complementary thrusts: measurement strategies that can amplify our coverage by orders of magnitude; innovations in IPv6-specific algorithms to infer router-level topologies; and analysis and remediation of security and privacy risks that our measurements reveal. Our goals are directly responsive to the CNS:Core's solicitation of "comprehensive, pervasive, accurate, and usable measurement capabilities", focused on *mapping the IPv6 Internet*.

- **Task A: New Measurement Strategies To Meet IPv6-specific Challenges:** We recently introduced a new active measurement technique, Yarrp [11], and demonstrated [15] its ability to discover >1.3M unique IPv6 router interfaces from a single vantage in a single day, an order of magnitude more topology than state-of-the-art IPv6 mapping systems using hundreds of vantage points. To operationalize this capability for use in scientific research, we need to refine the technique to support distributed deployment across hundreds of points, decouple probing from response collection, incorporate our previous work on intelligently directing IPv4 probing [10, 13], and improve efficiency (e.g. topological yield measured in interfaces or routers per unit time, per trace, or per packet) and performance (e.g. metrics of speed,

recall, per-hop responsiveness, reachability, network load, host load, minimizing complaints, etc). We also propose to map the IPv6 infrastructure of mobile providers, a reported driver of IPv6 adoption [58, 5, 64].

- **Task B: Innovations in IPv6-specific algorithms to infer router-level topologies:** Our previous collaboration yielded the first IPv6 address alias resolution technique [14, 43], which reduces the interface-level graph to a router-level graph. Unfortunately, due to router implementation differences and network filtering policies, our technique only works for a subset of all deployed routers (approximately one-third of probed interfaces [43]).[1] We will explore the use of multiple alias resolution techniques, e.g. [51], at Internet scale, as well as develop new IPv6 alias resolution techniques that leverage DNS hostname conventions to infer aliases. We will also develop methods to generate *candidate* aliases to discover previously unknown interfaces, which will facilitate comparative graph analyses between IPv4 and IPv6 router-level topologies, illuminating the degree of inter-dependence and shared risks – or lack thereof – between IPv4 and IPv6 infrastructure.
- **Task C: Analysis and Remediation of Security and Privacy Risks:** The most disturbing discovery thus far from our IPv6 active topology probing is pervasive use of EUI-64 addresses [34], which encode the interface's hardware MAC address into the low-order 64 bits of the IPv6 address as a mechanism to provide uniqueness. To mitigate the risks of using a static hardware identifier at the network layer (thus visible to a remote adversary), modern operating systems have adopted *privacy extensions* that utilize pseudo-random and ephemeral addresses [49]. Thus, we were surprised that approximately 45% of router hop addresses found in our initial Yarrp6 high-speed survey were EUI-64 addresses, assigned to Customer Premise Equipment (CPE) routers in large, homogeneous IPv6 deployments. The use of EUI-64 addresses to number network infrastructure triggers immediate concerns. First, MAC addresses leak information not only about the device manufacturer, but also potentially the model [47]. In these cases, attackers can perform targeted attacks, particularly devastating for generally ill-maintained consumer-grade equipment. Further, CPE using EUI-64 addresses can compromise privacy-focused efforts to circumvent IP-based tracking [50]. We plan to not only investigate this phenomenon, but also initiate efforts to resolve the vulnerability among providers, manufacturers, and work within standards bodies.

2 Motivation

Progress along these three fronts can help provide a solid foundation for networking and systems research that depends on network topologies. More broadly, the resulting data can enable not only researchers, but also policy makers and regulators to better understand the network. In this section, we provide specific benefits our proposed IPv6 topology research will provide for: i) resolving the IPv6 adoption paradox; ii) science and society; and iii) engineering and operational network practice.

2.1 Resolving the Adoption Paradox

Measuring IPv6 adoption is not straightforward [21]. Some proposed metrics of adoption include the number of networks (e.g. ASes) advertising IPv6 BGP prefixes, the fraction of content web sites with IPv6 DNS records, counts of client IPv6 web accesses, or the volume of IPv6 traffic on a network link. Further complicating accurate IPv6 adoption estimations are variations due to methodology and observation point.

In many countries, mobile networks are reportedly driving IPv6 adoption and traffic [58]. Apple has required all iOS apps to work in a IPv6-only world since June 2016 [7]. T-Mobile has even presented on their efforts to turn off IPv4 entirely [17]. Verizon Wireless claims to be one of the largest production IPv6 deployments, with a large IPv6 user base (handsets) [64]. In fact, Akamai reports that Verizon is the largest contributor of IPv6 requests to their CDN platform, estimating Verizon's IPv6 "adoption" rate at 83.6% [5], with no definition of "adoption." Similarly, Akamai's academic publications claim "a striking volume of World-Wide activity on IPv6 today." Yet, these reports stand in contrast to published measurements of

[1] Notably, speedtrap does not work on Juniper routers.

observable IPv6 traffic volume: approximately 1% IPv6 traffic in 2018 on three U.S. backbone links [33]; 2-3% on AMS-IX [1]. One benefit of improved IPv6 topology maps that reflect actual *reachability* is to better reconcile these wildly divergent reports of IPv6 penetration.

Rather than relying on aggregate and imprecise statistics used for marketing, we require systematic, broad, and representative, i.e. scientific, IPv6 measurements. For example, one explanation for the disparity in adoption statistics could be that the IPv6 traffic is traveling between directly connected private networks (e.g., Verizon and Facebook) and thus does not cross the public Internet. But the ways in which IPv6 is being deployed to mobile customers has important implications for users, and for the future of the Internet. Does IPv6 provide customers with unimpeded end-to-end connectivity? How much of the IPv6 Internet can successfully be reached? Are IPv6 paths within mobile networks shorter (either topologically or geographically) than over IPv4? What are the privacy and security implications of providing customers public IPv6 addresses? We can answer these questions by performing active topology mapping both toward mobile networks, as well as from and within mobile networks (§4.1). In the former case, we will leverage our high-speed probing infrastructure in conjunction with carefully constructed mobile network seed sets. To obtain traces from mobile networks, we will implement a mobile version of Yarrp, initially targeting the Android platform.

2.2 Benefits to Science, Engineering, and Society

Better tools and methods for IPv6 topology measurement will benefit not only scientific study of the Internet and other large scale networks, but also inform network engineering, security, and operational practices.

- **Third-party detection and localization of congestion:** As notions of network collaboration, competition, regulation, and neutrality continue to evolve, it is increasingly important for *independent* research to shed light on current practices and their implications. Network mapping is instrumental in locating not only where and how providers interconnect, but also when those connections are congested and impact consumers [44, 24]. Current approaches for such congestion mapping rely on exhaustive mapping of the IPv4 space, which is not feasible for IPv6.

- **Studying Censorship:** Network topology is often utilized to reverse engineer various forms of network-based censorship. Our prior work on understanding censorship-based outages during the Arab Spring relied on (IPv4) traceroute to understand how these methods evolve, and when they occur [23]. The much larger and more flexible address space of IPv6 amplifies the need for rapid, comprehensive IPv6 topology mapping during events of geo-political interest.

- **Understanding Routing Policies:** Active topology mapping reveals the actual data-plane path of traffic. For instance, when two ASes connect in a customer-to-provider relationship, we may expect very different traffic behavior than when two ASes have a settlement-free peering arrangement. Interdomain routing policies, although generally proprietary and non-public, are amenable to reverse engineering through use of topology mapping. Some routing policies require that data-plane traffic will only traverse particular paths and links contingent on conditions, e.g. during an outage when the policy-preferred path is down. In such cases, rapid topology mapping can expose the existence of the true, more complex topology. Further, while significant work has leveraged IPv4 topology data for AS relationship inference [45], and researchers have begun work on IPv6 AS relationship inference [31], little work has explored how IPv6 topology data can inform these inferences.

- **Graph theoretic analysis:** As a complex, dynamic, and organically growing graph, the structure of the Internet topology has been studied (and debated) for years [66]. Properties of Internet topology have implications for network survivability and resilience (e.g. if the degree distribution follows a power-law [66]) but also for the role of large network players and exchange points (e.g. if the topology

is flattening [30]). In this vein, the extent to which the IPv6 network mirrors the IPv4 topology, and whether the IPv6 and IPv4 topologies are *inter-dependent* remain important scientific questions.

2.3 Network Engineering

From a network engineering perspective, better IPv6 mapping is crucial to:

- **Detection and Mitigation of Route Hijacks:** A persistent problem within today's routing system is prefix hijacking [9]. A fundamental component to discriminating real hijacks from normal network events is visibility into both the control (BGP) and data plane [57]. As BGP hijack attacks shift to IPv6, accurate and fast IPv6 topology mapping is required to provide this data plane introspection.

- **Geolocation:** Mapping IP addresses to their physical location is an inexact science, yet is of critical importance to everything from content distribution, to attack attribution, fraud protection, and enforcing country-specific policy. Network topology plays an important role in informing IP geolocation [65, 39]. Our own exploratory work has shown that IPv6 geolocation is especially difficult given circuitous IPv6 routing, tunneling, and IPv6 address agility [61]. Rapid and more complete IPv6 topology mapping has the potential to significantly advance IPv6 geolocation.

3 Background and Related Work in IPv6 Topology Measurement

While decades of research have developed and refined active IPv4 topology discovery (e.g. [59, 38, 26, 40, 13]), none of this work addresses IPv6-specific mapping challenges. We briefly review prior work to establish context for our proposed research.

3.1 Measurement Methods

BGP-based measurement. Prior IPv6 topology analysis has largely avoided active probing due to the sheer size of the address space and the sparsity of infrastructure within it. Using passive BGP data, Dhamdhere et al. found fewer than 50% of AS-level paths in 2012 were identical between IPv4 and IPv6, with a single AS (Hurricane Electric) clearly dominant in the IPv6 topology [25]. In 2013, Czyz et al. found only 19% of IPv4 ASes also supporting IPv6, most of them large transit networks [21].

Traceroute-based measurement. Two production platforms continually perform active IPv6 topology measurement. CAIDA's Archipelago (Ark) [18] and RIPE Atlas [54]. These systems send Paris traceroute probes [8] toward the :: 1 address in each IPv6 prefix present in a global BGP table.[2] Using BGP prefixes to guide target selection captures topological breadth, but not depth, i.e. it fails to discover subnetting. Our 2014 study exhaustively probed an address in each routed /48 – 400M traces from 26 vantage points found more than double the number of unique interfaces compared the coarser granularity probing based on global routing tables [56]. Our recent development of the first IPv6 router alias resolution methods [14, 43] relied on these traceroute data sets, but did not solve the problem of creating input lists of IPv6 interfaces.

3.2 State of the Art in High-Speed Topology Probing

Current active Internet topology discovery techniques require significant probing time, during which the network may change or experience transient dynamics, rendering the notion of a global network topology "snapshot" misleading [11, 13]. Toward a more comprehensive and accurate measurement capability, we developed Yarrp (Yelling at Random Routers Progressively), a novel technique to permit high-speed, Internet-scale active topology probing [11]. Essentially, Yarrp is a high-speed version of traceroute that scales to Internet-wide probing[3]. Yarrp overcomes two primary impediments to the speed and scale of conventional approaches: i) per-trace state maintenance; and ii) a low-degree of parallelism. Yarrp is stateless, reconstituting

[2] CAIDA's Ark probing relies on scamper, the current state-of-the-art topology probing tool, which supports Paris traceroute [42].

[3] In contrast, traceroute was never designed for Internet-wide probing or gathering complete Internet topologies, but rather as a diagnostic tool for network operators to probe individual paths.

all necessary information from ICMP replies as they arrive asynchronously. To avoid overloading routers or links with probe traffic, Yarrp randomly permutes an input $IP \times TTL$ space (probing different portions of many different paths simultaneously). This means, for example, that Yarrp may send a probe to IP address A with $TTL = 12$, then B with $TTL = 3$, then C with $TTL = 9$, and so on until the entire space of $TTLs$ for each target IP is covered. In experimental deployment, Yarrp probed at more than $200kpps$ and discovered more than 500,000 IPv4 router interfaces in under 30 minutes from a single vantage point.

As alluded to in the introduction, IPv6 presents unique challenges for topology mapping. One of those challenges is mandated ICMPv6 rate-limiting, which is generally implemented on routers using a token bucket [20]. Figure 1 illustrates the implications of ICMPv6 rate-limiting on conventional traceroutes, i.e. those that send probes with monotonically increasing TTLs starting at 1. The graph plots the per-trace fraction of responses received versus TTL of the probe at speeds of 20, 1000, and 2000 pps toward CAIDA's April 2018 set of IPv6 targets [18].

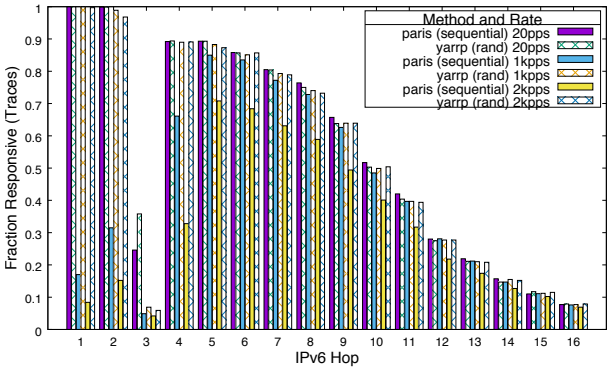
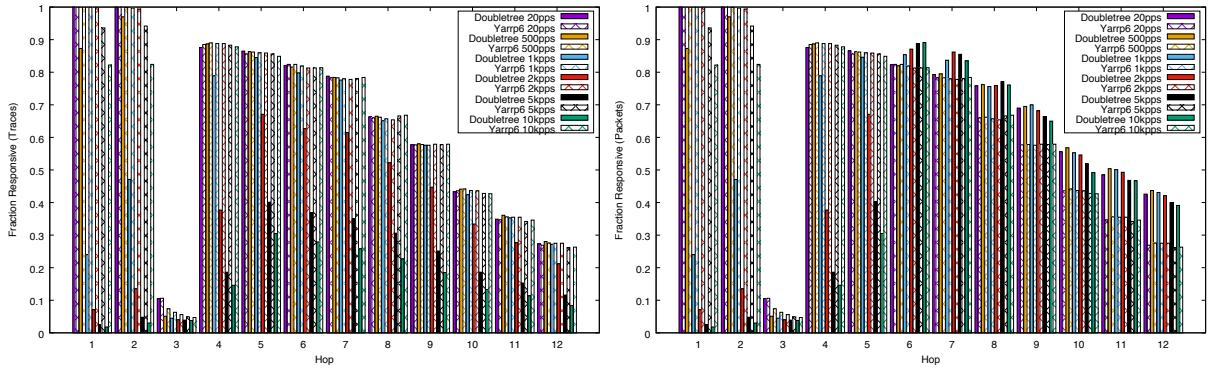


Figure 1: Probing strategy and rate versus responsiveness. ICMP6 rate-limiting is evident, as well as per-hop differences. Yarrp6's randomly permuted probing order improves per-trace responsiveness.

We used two probing strategies: randomized (as implemented in Yarrp6) and sequential (as implemented in scamper [42]). The number of responsive interfaces decreases as hop distance increases, so we compare relative performance at different speeds. While Paris and Yarrp6 have nearly identical response rates at 20pps (Figure 1) Yarrp6 yields a 100% response rate from the first hop for 1000 and 2000pps as compared to under 20% and 10% for sequential traceroute. Across both vantage points, at higher probing rates we observe better performance with Yarrp6 at all hops.

Optimizing Probing Order to Minimize Effects of Rate Limiting: In 2005, in an effort to improve traceroute efficiency, Donnet et al. developed Doubletree. Doubletree is a technique to avoid re-probing the same initial segments of many traceroute-observed paths from the same vantage points [26]. Doubletree chooses an intermediate starting TTL and probes forward (increasing TTL) and backward (decreasing TTL) until it receives a response from an interface it has previously observed. Doubletree then interpolates missing portions of the path based on previous results. Akamai researchers leveraged these ideas to avoid rate-limiting for IPv6 probing [6].

To gain intuition over Doubletree's performance relative to Yarrp6 as influenced by ICMPv6 rate limiting, we recently probed the same April 2018 CAIDA target set [18] from one of our vantage points at various packet rates using scamper's Doubletree implementation. Figure 2 displays the per-trace and per-packet response results as a function of router hop and probing speed. While Doubletree induces less rate-limiting than traditional traceroute methods, we observe an unexpected effect: when rate limiting causes a hop to be non-responsive, Doubletree continues probing backward. Thus, as the token buckets on the initial hops drain, Doubletree continues to probe these initial hops, keeping their buckets empty. Thus, rather than having the intended effect of minimizing redundant probing on the initial common portion of the traceroute path, ICMPv6 rate-limiting can cause Doubletree to devolve to the same lossy behavior we observe with sequential traceroute. Notwithstanding this behavior, Doubletree has two fundamental limitations. First, it must heuristically select the intermediate starting TTL for each vantage point. Second, using previous measurements to fill gaps in paths can induce erroneous path inferences [28]. We believe Yarrp6 offers a better compromise between completeness and scalability, better-suited to Internet-wide IPv6 topology studies. To date, Yarrp has not been deployed in production.



(a) Doubletree (per-trace)

(b) Doubletree (per-packet)

Figure 2: Toward understanding efficiency: comparing Doubletree [26] and Yarrp6 [15]. While Yarrp6 outperforms Doubletree on a per-trace basis, Doubletree performs better at higher TTLs simply due to its stateful nature. Combining Yarrp techniques with stopping heuristics may thus provide an ideal approach.

Maximizing Probing Efficiency: In addition to the need to maximize topological discovery, the size of the IPv6 space implies an objective of maximizing efficiency, i.e. capturing new interfaces or routers per unit time, or per trace, or per packet. While Yarrp address the problem of how to probe, we have also examined what to probe and evaluated several hitlists and destination selection strategies [15] (Even with Yarrp's speed, exhaustively probing the IPv6 space is infeasible). Figure 3 shows the relationship between probe count and unique IPv6 interfaces discovered for five different strategies. As a baseline, the CAIDA strategy represents the current production method of selecting the base `::1` address in each routed prefix [18]. As a second baseline of BGP-based target selection, we probed randomly generated (routable) IPv6 addresses. To explore the power from using recently derived IPv6 hitlists, we probed the Fiebig hitlist [27] and used 6gen [48] to generate likely IPv6 addresses from input router seeds. Finally, "cdn-k32" consists of targets derived from anonymous aggregates of web clients observed at a large CDN [53].

The base `::1` probing strategy performs best in initial stages of the probing, but suffers a noticeable flattening past 300k packets. Further, it peaks at fewer than 100k interfaces after ∼2M probes as it exhausts the target set. This dichotomy of performing well initially, but falling short of the absolute number of interfaces discovered via other strategies illustrates how this naïve strategy lacks the specificity to discover the significant topology that exists in IPv6 subnetting.

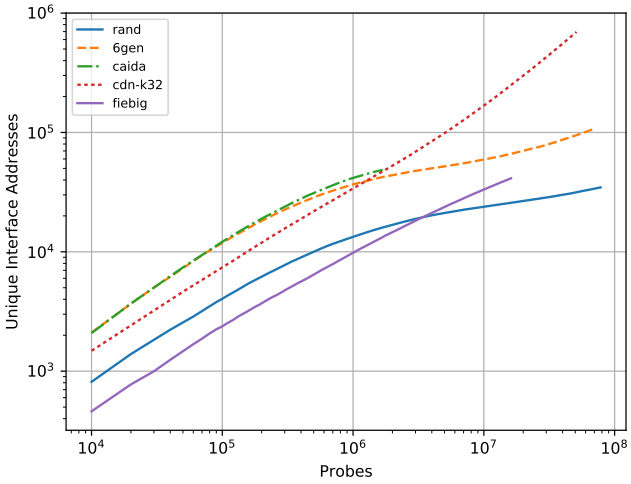


Figure 3: Evaluating probing strategies: discovery vs. probe packets (log-log). Intelligent probing improves discovery.

Unsurprisingly, random, unguided target selection performs poorly, with a precipitous drop in newly discovered interfaces after ∼1M probes. However, random outperforms Fiebig's hitlist prior to this point, largely due to the high degree of clustering (many targets are within the same `/64` network) present in the Fiebig hitlist. Similarly, the 6gen strategy provides a high interface yield at the onset of probing, but flattens past 1M probes. In fact, the shape of the 6gen curve closely mirrors random, but with a fixed positive offset. In contrast, the cdn-k32 performs best, continuing to yield topological results over the lifetime of the trace, largely because its ability to discover both topological breadth as well as depth from subnetting [15].

4 Research Plan

4.1 Task A: New Measurement Strategies To Meet IPv6-specific Challenges

Our first task is to advance the state-of-the-art in IPv6 measurement to provide, deploy, and apply the most "comprehensive, pervasive, accurate, and usable measurement capabilities". This task includes techniques to improve efficiency and performance; enabling IPv6 Internet topology mapping from distributed VPs; decoupling probing from response collection; and adaptation of our work on intelligently directing IPv4 probing [10, 13]

Techniques to improve efficiency and performance: Currently, Yarrp provides a simple approach to large-scale scanning without the complexity of intelligently driving the probing. This approach favors *recall* – finding the most infrastructure possible in the shortest amount of time – over efficiency. Yarrp's statelessness implies that it does not stop probing after reaching a destination or after not receiving replies from several hops in sequence (the "gap limit"). Figure 2b illustrates the limitation of this simple approach: Yarrp6 outperforms Doubletree at TTL hops below the configured Doubletree midpoint, but as TTL increases, Yarrp6 yields lower per-packet responsiveness. In short, Doubletree stops probing after the gap limit or after reaching the destination, but Yarrp continues probing, past the point where it will likely receive a response.

We propose to experiment with our recently introduced idea of a "fill mode," [15] and rigorously investigate its use, performance, or efficiency gains. In fill mode, if the Yarrp6 listener receives a response for a probe sent with hop limit h that exceeds the max probing TTL, Yarrp6 immediately sends a new probe toward the destination with a hop limit of $h + 1$ to fill the gap. While these additional probes are not randomized, fills are uncommon, and occur at the tail of the path where the effect of sequential probing has the least impact. We are especially interested in cases where the fill mode prematurely stops, for instance when there is an unreachable hop past the configured TTL m, but a subsequent reachable hop at $m + k$ that is not probed. While maintaining a count of unreachable hops is not possible in a stateless implementation, we plan to study mechanisms to enhance probing efficiency. This presents an opportunity for synergy if we can merge elements of Doubletree into Yarrp.

Parallelized Probing: The sheer size of the IPv6 address space implies that mapping speed will always be a limiting factor in comprehensive probing. Parallel execution of the Yarrp process on distinct vantage points will provide the most significant practical speed improvement. A significant advantage to Yarrp's design is the potential to easily randomize and distribute the probing to multiple vantage points with negligible coordination and communication overhead.

However, distributed Yarrp is an open research problem. While the entire domain of targets and TTLs can be subdivided among vantage points, we wish to ensure that different vantages are not responsible for different portions of the TTL space for the same target. Otherwise, the discovered hops toward a destination will not reflect the true path, but rather interspersed portions of paths. A hypothetical solution is a simple scheme that assigns each of n vantage points a unique identifier. If each vantage point uses the same key, it will generate the same randomly permuted sequence of targets and TTLs. As each vantage point works through its locally computed sequence, it checks whether it is responsible for the target via a simple modulo computation (e.g. $IP\%(n-1) \stackrel{?}{=} vp$). For additional randomness, the IP may be hashed prior to this check.

But this approach wastes CPU cycles (and hence reduces Yarrp's probing speed) as Yarrp instances skip portions of the permutation space they are not responsible for. We therefore propose to investigate more elegant solutions to distributing Yarrp. To explore techniques and research toward parallelizing and distributing Yarrp, we will deploy Yarrp on CAIDA's Ark platform. Ark can help inform important systems questions that are required to produce the best possible topology maps, including determining the best way to retrieve and aggregate data from the distributed monitors, and how to manage vantage points with varying probing rates. Finally, we will validate and run Yarrp in a distributed fashion on the live Internet.

Decoupling Probing and Receiving: Yarrp's stateless nature opens the possibility to decouple Yarrp's

probing engine from collection. Because Yarrp recovers state from information contained within the probe responses, the machine that receives probe responses does not have to be the one that sends the probes. To expand measurement coverage, the Yarrp prober may put the source address of a remote collector machine inside probes, assuming common ownership of the machines or permission to spoof the source IPv6 address. CAIDA maintains vantage points with permission and capability to spoof for measurement research, so we can experiment with this capability. We hope to explore paths that currently block responses, reveal AS-internal infrastructure, and gain additional visibility into IPv6 exchange point connectivity.

Intelligent target selection: The studies described in Section 3.2 have advanced our understanding of how to identify address structure and predict likely portions of active IPv6 address space, e.g. as targets of IPv6 measurement surveys. However, they have focused on IPv6 end-hosts rather than core network infrastructure or finding targets for active topology measurements. Our experiments with 6gen [15] generated many permutations of likely hosts within single /64s. This subtask will adapt our previous work on intelligently directing IPv4 probing [10, 13] to our Yarrp system. We will use similar techniques to discover structure within the high-order bits, or network portion, of the router infrastructure address we discover from prior rounds of probing to drive the selection of targets to probe in subsequent rounds. Open research issues in target selection include how to best derive a single set of targets for production deployment, how to utilize hitlists where addresses within the hitlist change rapidly over time, and, more generally, how to improve target selection to maintain efficiency and coverage (high per-probe recall).

Mapping Mobile Providers: As described in the motivation, mobile providers have claimed large IPv6 infrastructure deployments and large numbers of IPv6 customers. Similarly, content providers have claimed large numbers of IPv6 requests, clients, and "striking" volumes of IPv6 traffic. To address the larger question of resolving these adoption claims with other reported metrics that show lower IPv6 usage, we propose a two-pronged effort to map the IPv6 infrastructure of mobile providers.

First, we will identify IPv6 prefixes belonging to mobile providers and delineate between IPv6 prefixes used to number mobile infrastructure versus IPv6 prefixes that are assigned to customers. While seemingly simple, automatic (i.e. scalable) labeling of prefixes as either mobile or not is an inference problem that has not been solved. We propose to utilize an analysis of HTTP user-agents observed on backbone links and other passive observation points to find prefixes that are predominantly used by devices with a mobile user-agent string. Since a mobile user-agent string can as easily originate from a residential broadband address when the device is on WiFi, we will build a statistical model of user-agents per prefix to identify regions of mobile IPv6 user addresses. Further, we will leverage the proposed DNS research in Task B (§4.2) to see if hostname analytics can help.

Second, we will utilize Yarrp to perform high-speed mapping of identified mobile networks. Yarrp's speed comes into play in two ways here: to cover as much of the provider's address space as possible, and to identify ways that providers employ ephemeral address and address agility. Finally, we propose to utilize general purpose computers tethered to mobile devices to serve as the source of Yarrp probes originating *from* the mobile edge.[4] Our initial experiments from Yarrp measurements from two U.S. mobile carriers have already yielded some fascinating insights, including different IPv6 paths than IPv4, shorter IPv6 paths as compared to IPv4, and even private and link-local IPv6 addresses along paths.

4.2 Task B: innovations in IPv6-specific algorithms to infer router-level topologies

Alias resolution – the problem of determining whether two or more interface address belong to the same physical router – is crucial to producing router-level topologies. Router-level topologies are frequently more useful and representative than raw IP interface graphs, especially when considering network structure and resilience.

[4] In contrast to IPv4, we do not observe NAT used for IPv6 when tethering.

Integration of existing and new methods to improve alias resolution completeness and accuracy. Our tool, speedtrap [43], is currently the best available method to resolve IPv6 aliases; it works by inducing IPv6 routers to send fragmented packets, and comparing the fragment identifiers between probed interfaces. Thus, speedtrap only works on routers that respond with ICMPv6 Packet-Too-Big messages, and if such packets are not filtered along the path, and only when routers implement particular IPv6 fragmentation behavior in their control plane. For instance, we find that while speedtrap works well for Cisco devices, it fails with Juniper routers, which implement random fragment identifiers. We propose to explore new methods of resolving IPv6 aliases, including combining speedtrap with techniques developed by other groups, including UAv6 [51] (whose author begins a postdoc at CAIDA in late 2018).

DNS naming convention inference to support topology mapping. We propose to leverage hostnames and DNS PTR records corresponding to interfaces of known aliased routers to infer router naming conventions. We will use statistical learning primitives to automate inference of per-provider regular expressions (REs) that capture a given provider's method of assigning names to router interfaces. We can also leverage this DNS-based approach to generate *candidate* aliases that we can subsequently probe with other alias resolution techniques. Our collaboration with Matthew Luckie at U. Waikato on this effort thus far suggests the promise of this approach, which significantly generalizes manual heuristic approaches to understanding names, e.g. as pioneered in [59]. We describe the intuition behind our idea as follows.

First, given a set of regular expressions (REs) that capture the per-provider naming convention, we will use these REs to generate candidate router names. As a proof-of-concept, we plan to permute the space of integers within known names, and then determine whether a corresponding `AAAA` (DNS) record exists for each candidate name. We will use active probing methods to determine if the candidate names with DNS records are new aliases of routers we have already identified, or an interface of a previously unknown router. To illustrate, consider the following interfaces discovered through CAIDA's IPv6 traceroute campaigns [18]:

```
xe-0-1-0-0-ce-bothwaak14w.bothell.wa.seattle.comcast.net
xe-1-1-0-0-ce-bothwaak14w.bothell.wa.seattle.comcast.net
xe-4-0-0-0-ce-bothwaak14w.bothell.wa.seattle.comcast.net
```

Given these names as input, we can expand the four interface identifier digits ("xe-X-X-X-X") to generate all possible 10^4 permutations. The intuition here is that the digits typically signify facilities, routers, router line cards, and router interface numbers. For instance, "xe-2-3-0-0" might denote the second router, third line card, and first interface on the line card in the first data center. In the above example, after generating these permutations, we discover that:

```
xe-1-0-0-0-ce-bothwaak14w.bothell.wa.seattle.comcast.net
```

has an authoritative DNS record, and alias resolution confirms that it is an alias of the three other interfaces. Our initial experiments with this idea have convinced us that not only can the DNS be used to associate previously unassociated interfaces, it can also be used to discover *entirely new interfaces and aliases*, i.e. those not previously known via the traceroute campaign.

Similarly, we will use other existing list of DNS names, e.g., gathered via walking reverse DNS zones [27], to test candidates for alias resolution.

Finding IPv4/IPv6 Aliases. In an IPv6 world, aliases extend beyond identifying IPv6 interface addresses that belong to the same router to the problem of inferring whether an IPv4 address and an IPv6 address are aliases, i.e. whether they are assigned to the same interface. Such inference is crucial to understanding the extent to which the IPv6 infrastructure depends (or does not) on the IPv4 infrastructure, and whether these networks share fate e.g. during attacks. Our previous work used TCP timestamp skew to find server *siblings*: IPv4 and IPv6 addresses that belong to the same physical server [12]. We inferred cross-protocol siblings by using the common upper layer, namely TCP, and included a measurement study against DNS and web servers.

Our technique leveraged the well-known observation that the clock skew of machines, as exposed remotely by TCP timestamps, can be used to fingerprint physical machines. However, routers typically do not expose open TCP ports, preventing use of this method to find sibling addresses on routers.[5] But we can use insights from these studies to develop novel methods for finding IPv4 to IPv6 siblings. First, we will investigate the use of DNS regular expressions (described above) to find alias candidates. We will also investigate other identifiers, including MPLS labels [62], and TTL-based router signatures [63] to find cross-protocol aliases.

Topology Analysis. The combined effort of our next-generation IPv6 topology probing, alias resolution, and alias discovery methods, along with reliable production deployment (see §4.2) will allow us to produce the most complete IPv6 topologies currently available to researchers. Once we have our new measurement capabilities deployed, we will analyze the resulting data sets by comparing IPv6 snapshots over time, and comparing IPv4 and IPv6 topologies.

Comparing IPv6 Topology Snapshots. Because Yarrp probes links in a randomly permuted, responses return in a similarly random order. This approach inherently decouples probing operation from path reconstruction – Yarrp can only reconstruct the full path to any single destination after all probing completes, for all destinations. This path reconstruction can require significant computation, but can occur offline and is not time critical. This decoupling means probing can be very fast, which brings us closer to the goal of obtaining true topology snapshots, where we can compare differences to capture fine-grained topology dynamics. We propose such comparisons of IPv6 network topologies to better understand not only the IPv6 network's behavior, but also help shed light on e.g. short-lived route hijacking events, discover backup links that are only exposed during transient outages, inform IP geolocation, and gain rapid insight into geo-political events such as censorship.

Comparing v4/v6 Topologies. Perhaps the most interesting analysis will leverage our expertise in collecting and collating IPv4 topologies to perform comparative analyses of the IPv4 and IPv6 logical networks. While prior work has compared IPv4 and IPv6 at the AS-level through passive BGP analysis, accurate IPv6 topologies would enable the first comparative studies of these two networks at the interface and router-level. We will not only be able to compare the structural characteristics of the IPv4 and IPv6 topologies, but also to analyze the extent to which they share physical infrastructure, and hence fate. As the IPv4 and IPv6 networks converge, we will quantify the extent to which not only attacks, but disruptions and bugs, that affect one one infrastructure will impact the availability of the other.

Production Deployment: A long-standing initiative of CAIDA has been to not only perform research, but provide infrastructure and resources to perform long-running continuous measurements. The data from these production measurement campaigns has proved invaluable – both for analyzing longitudinal trends and understanding how properties of the network evolve, but also for retroactively analyzing particular important events, e.g. attacks, censorship, or peering changes, e.g. [23, 44]. We will incorporate the research results of Task A into our publicly available Yarrp tool [11, 15], and perform the necessary systems development to bring Yarrp into a production state on CAIDA's Archipelago platform [37]. A new production IPv6 topology measurement system represents a significant systems development effort including: making the tools robust, intelligently distributing measurement tasks among a large pool of (currently more than 200) vantage points, handling much higher probing rates (10× more packets/sec), and processing and maintaining much larger volumes of topology data. These development efforts are essential to operationalizing a sustainable system that can provide invaluable continuous IPv6 topology data to serve as a foundational component supporting future network research.

[5] Although Czyz et al. find approximately 4% of 25k dual-stack routers respond on the ssh and BGP ports [22].

4.3 Task C: Analysis and Remediation of Security and Privacy Risks

An unexpected finding of our initial IPv6 active topology discovery campaign is non-trivial numbers of EUI-64 addresses [4], which encode the interface's hardware 48-bit IEEE Media Access Control (MAC) address [2] into the low-order 64 bits of the IPv6 address as a mechanism to ensure uniqueness [34]. While this standard provides a simple mechanism for hosts to automatically configure and use a unique IPv6 address on a subnet [60], it effectively leaks the hardware MAC address of the host to higher layers of the network stack. Typically, MAC addresses are only visible to passive observers in the broadcast domain of the same layer-2 subnet. But any layer-3 device along the communication path can observe these EUI-64 IPv6 addresses – and, in turn, observe the host's MAC address. These addresses are static, associated with the physical device, and if known, both leak device details and act as a unique tracking ID.

Modern operating systems have evolved to mitigate this risk by randomizing the host bits of their IPv6 addresses according to the privacy extensions standard [49]. Not only are these addresses randomized, they are highly ephemeral [52]. While end-hosts have thus largely abandoned the use of EUI-64 addresses, Customer Premise Equipment (CPE) infrastructure (e.g. home routers and gateways) still extensively use such addresses. Traceroute probing reveals these addresses, introducing two vulnerabilities: i) the ability to identify the manufacturer and model of a device, thereby permitting targeted attacks [46, 47]; and ii) the ability to track users despite efforts to prevent such tracking. More specifically, not only are the host bits of a client's IPv6 address, but their entire assigned network prefix, is similarly ephemeral [50]. Many ISPs regularly change client IPv4 and IPv6 addresses for privacy reasons [3]. However, while the client's end-host IPv6 address is changing, the address of her home gateway is not, so an attacker can perform traceroutes to the client can determine that it is the same client, and track the client's assigned addresses over time. We will take a three-step process to investigating the prevalence and risk of EUI-64 addresses in IPv6 *infrastructure*.

Longitudinal Characterization: Our initial IPv6 topology survey [15] produced responses from 651.4k unique EUI-64 interface addresses, or 45% of all hops. Of these, 59% were from one of just two manufacturers, and 99.9% of those devices were in just two ISP networks, each in different countries. In both cases, WWW content suggests they were Customer Premises Equipment (CPE) routers in large IPv6 deployments.

Our first sub-task is longitudinal characterization of EUI-64 SLAAC addresses discovered through operational probing by three Internet wide measurement platforms: Ark [18], Atlas [55], and Yarrp6 [16]. We will study the evolution and prevalence of EUI-64 addresses, as well as classify them by networks, providers, and geo-regions hosting them. Our initial analysis of CAIDA IPv6 topology data in 2018 [18] suggests that infrastructure EUI-64 addresses are pervasive: 70% of observed ASes have them, 20% have 25 or more. CAIDA's probing of the base address (`::1`) for this data imply these numbers are a lower-bound; the true number of EUI-64 addresses deployed is likely much higher, which targeted high-speed Yarrp6 probing can help reveal.

The distribution of MAC addresses among the prefixes of different providers exhibits oddities that bear further investigation. For instance, particular MAC addresses are disproportionately represented in some networks and countries, suggesting the existence of multiple devices with the same MAC address. Such MAC address duplication has been anecdotally reported on in the past, but not formally studied. Addresses may be duplicated by manufacturers of inexpensive equipment that do not wish to pay for new IEEE allocations, or may be reused after the expected useful lifetime of the hardware has passed. We hope to shed light on this practice.

Given a set of passively observed EUI-64 addresses, we will perform active measurements toward combinations of the provider's prefixes and the EUI-64 address, searching for the client's CPE within the provider's address space. Further, many EUI-64 addresses concentrate among subnets of globally advertised prefixes (i.e. more specific than the advertised prefix). We believe that high-speed tools such as Yarrp6 can efficiently, effectively, and exhaustively probe these subnets in order to discover previously unknown CPE infrastructure. Proof-of-concept experiments thus far have been encouraging: random scanning of the /64

prefix containing the most EUI-64 addresses in a month's worth of passive CAIDA data [18] resulted in 18 responses from EUI-64 IPv6 addresses, where the MAC addresses were previously unobserved.

4.4 Granular Mapping of CPE MAC Addresses to Hardware

MAC addresses reveal information about the specific device model [47]. The high-order three bytes of a MAC address, termed the Organization Unique Identifier (OUI), are allocated to a manufacturer, for use as they wish. Given samples of known devices and their MAC addresses, we can build a map of likely allocations, and use it to predict the device model given an unknown MAC address. In prior work [47], we created such a mapping for mobile devices since a passive observer can observe MAC addresses of 802.11 WiFi networks. Allocations of contiguous blocks to distinct device models are also present among access point (AP) manufacturers. Figure 4 shows the inferred allocation of the `C0:C1:C0` OUI owned by Cisco, with 17 models of APs. We find 248 distinct contiguous ranges dispersed throughout the OUI; the Linksys E1000 constitutes the largest contiguous block of any single device (2.4M addresses). About 60% of the OUI was allocated, with a mean of ∼41K addresses per block.

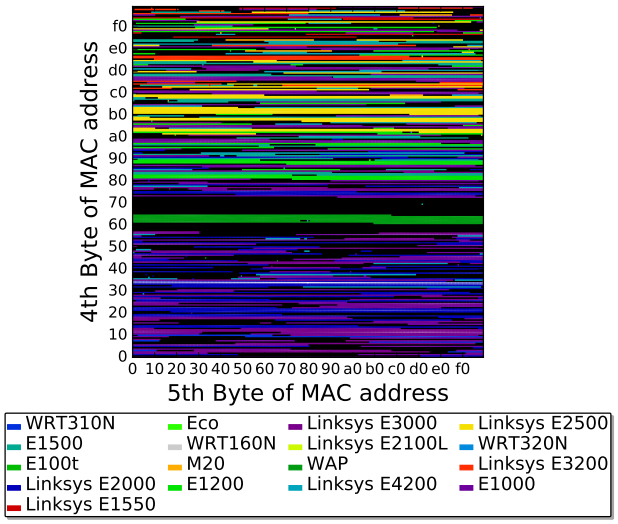


Figure 4: Observed Models in `C0:C1:C0` (Cisco)

Remote discovery of IPv6 infrastructure EUI-64 allows creation of analogous fine-grained mappings of MAC addresses to model for core network *infrastructure*. The ability to determine the specific model of a piece of network equipment, in addition to the manufacturer, poses a security vulnerability, namely enabling targeted attacks [19, 41]. We propose to first research the scope of the problem by building the first comprehensive mapping of CPE MAC addresses to fine-grained manufacturer and model.

Key to our ability to produce accurate mappings is having ground-truth for specific OUIs, i.e. we must know the specific device manufacturer and model corresponding to a sample of MAC addresses within the space. Many ground-truth examples are available on web sites, email lists, and discussion forums, including by homeowners having trouble configuring CPEs, and manufacturers trying to help them. We will use basic search engines to automate discovery of these MAC addresses and their corresponding device models where we can, and manually populate data otherwise. We will use the various search engine APIs to issue queries for a particular device model and mine the results for MAC addresses. Since MAC addresses have a distinct format, we can discover likely addresses corresponding to that model.

4.5 Remediation

Finally, we will demonstrate the potential to exploit this information, as well as understand the scope of the security and privacy impacts. We will run on-demand traceroutes to IPv6 clients that access the web and find those that, over time, have a common last hop address with an EUI-64 address. To perform this experiment, we will use CAIDA's web server and partner with other organizations to expand coverage, e.g. we have partnered with Akamai in the past, and made use of their large web footprint in other research.

We believe that these leaked MAC addresses via EUI-64 number of infrastructure represent a currently unknown security vulnerability, and it is our responsibility to responsibly disclose the vulnerability and work with providers and networks at remediation. The community should carefully consider the implications of these router-address-assignment practices as a router's source address, alone, e.g. when sending ICMPv6 error messages, can disclose private details. We will engage with our network of engineers and network

managers that we have cultivated over the years, as well as present our findings to equipment manufacturers, NANOG, and the IETF.

References

[1] AMS-IX IPv6 traffic statistics. https://ams-ix.net/technical/statistics/sflow-stats/ether-type.

[2] Ieee standard for information technology - telecommunications and information exchange between systems - local and metropolitan area networks - specific requirements part 3: Carrier sense multiple access with collision detection (csma/cd) access method and physical layer specifications. *IEEE Std 802.3-2005 (Revision of IEEE Std 802.3-2002 including all approved amendments)*, pages 1–2695, Dec 2005.

[3] Zwangstrennung (forced ip address change), 2018. https://de.wikipedia.org/wiki/Zwangstrennung.

[4] D. E. 3rd and J. Abley. IANA Considerations and IETF Protocol and Documentation Usage for IEEE 802 Parameters. RFC 7042 (Best Current Practice), Oct. 2013.

[5] Akamai. State of the internet, 2018. http://www.stateoftheinternet.com/ipv6.

[6] P. Alvarez, F. Oprea, and J. Rula. Rate-limiting of ipv6 traceroutes is widespread: measurements and mitigations., 2017. https://www.ietf.org/proceedings/99/slides/slides-99-maprg-rate-limiting-of-ipv6-traceroutes-is-widespread-measurements-and-mitigations-02.pdf.

[7] Apple. Supporting ipv6-only networks, 2016. https://developer.apple.com/support/ipv6/.

[8] B. Augustin, X. Cuvellier, B. Orgogozo, F. Viger, T. Friedman, M. Latapy, C. Magnien, and R. Teixeira. Avoiding traceroute anomalies with Paris traceroute. In *Proceedings of ACM IMC*, 2006.

[9] H. Ballani, P. Francis, and X. Zhang. A study of prefix hijacking and interception in the internet. In *ACM SIGCOMM Computer Communication Review*, volume 37, pages 265–276, 2007.

[10] G. Baltra, R. Beverly, and G. G. Xie. Ingress Point Spreading: A New Primitive for Adaptive Active Network Mapping. In *Proceedings of the 15th Conference on Passive and Active Network Measurement (PAM)*, volume 8362, pages 56–66, Mar. 2014.

[11] R. Beverly. Yarrp'ing the Internet: Randomized High-Speed Active Topology Discovery. In *Proceedings of the ACM SIGCOMM Internet Measurement Conference (IMC)*, Nov. 2016.

[12] R. Beverly and A. Berger. Server Siblings: Identifying Shared IPv4/IPv6 Infrastructure via Active Fingerprinting. In *Proceedings of the 16th Conference on Passive and Active Network Measurement (PAM)*, Mar. 2015.

[13] R. Beverly, A. Berger, and G. G. Xie. Primitives for active Internet topology mapping: toward high-frequency characterization. In *Proceedings of the ACM 10th annual conference on Internet measurement (IMC)*, pages 165–171, 2010.

[14] R. Beverly, W. Brinkmeyer, M. Luckie, and J. P. Rohrer. IPv6 Alias Resolution via Induced Fragmentation. In *Proceedings of the 14th Conference on Passive and Active Network Measurement (PAM)*, volume 7799, pages 155–165, Mar. 2013.

[15] R. Beverly, R. Durairajan, D. Plonka, and J. P. Rohrer. In the IP of the Beholder: Strategies for Active IPv6 Topology Discovery. In *Proceedings of the ACM SIGCOMM Internet Measurement Conference (IMC)*, Nov. 2018.

[16] R. Beverly and J. Rohrer. Yarrp6 Beholder Datasets, 2018. `https://www.cmand.org/yarrp/ipv6/`.

[17] C. Byrne. Breaking free of ipv4, 2014. `https://www.nanog.org/sites/default/files/wednesday_general_byrne_breakingfree_11.pdf`.

[18] CAIDA. The CAIDA UCSD IPv6 Topology Dataset, 2018. `http://www.caida.org/data/active/ipv6_allpref_topology_dataset.xml`.

[19] S. Checkoway, D. McCoy, B. Kantor, D. Anderson, H. Shacham, S. Savage, K. Koscher, A. Czeskis, F. Roesner, T. Kohno, et al. Comprehensive experimental analyses of automotive attack surfaces. In *USENIX Security Symposium*, 2011.

[20] A. Conta, S. Deering, and M. Gupta. Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification. RFC 4443, Mar. 2006.

[21] J. Czyz, M. Allman, J. Zhang, S. Iekel-Johnson, E. Osterweil, and M. Bailey. Measuring IPv6 Adoption. In *ACM SIGCOMM*, Aug. 2014.

[22] J. Czyz, M. Luckie, M. Allman, and M. Bailey. Don't Forget to Lock the Back Door! A Characterization of IPv6 Network Security Policy. In *Network and Distributed Systems Security (NDSS)*, Feb 2016.

[23] A. Dainotti, C. Squarcella, E. Aben, K. C. Claffy, M. Chiesa, M. Russo, and A. Pescapé. Analysis of country-wide internet outages caused by censorship. *IEEE/ACM Trans. Netw.*, 22(6):1964–1977, Dec. 2014.

[24] A. Dhamdhere, D. Clark, A. Gamero-Garrido, M. Luckie, R. Mok, G. Akiwate, K. Gogia, V. Bajpai, A. Snoeren, and k. claffy. Inferring Persistent Interdomain Congestion. In *ACM SIGCOMM*, Aug 2018.

[25] A. Dhamdhere, M. Luckie, B. Huffaker, k. claffy, A. Elmokashfi, and E. Aben. Measuring the Deployment of IPv6: Topology, Routing and Performance. In *Internet Measurement Conference (IMC)*, Nov 2012.

[26] B. Donnet, P. Raoult, T. Friedman, and M. Crovella. Efficient algorithms for large-scale topology discovery. *ACM SIGMETRICS Performance Evaluation Review*, 33(1), 2005.

[27] T. Fiebig, K. Borgolte, S. Hao, C. Kruegel, and G. Vigna. Something From Nothing (There): Collecting Global IPv6 Datasets From DNS. In *Proceedings of the 18th Passive and Active Measurement Conference*, Mar. 2017.

[28] T. Flach, E. Katz-Bassett, and R. Govindan. Quantifying violations of destination-based forwarding on the internet. In *Proceedings of the 2012 Internet Measurement Conference*, pages 265–272, 2012.

[29] O. Gasser, Q. Scheitle, Q. Lone, M. Korczynski, S. D. Strowes, L. Hendriks, and G. Carle. Clusters in the Expanse: Understanding and Unbiasing IPv6 Hitlists. In *Proceedings of the ACM SIGCOMM Internet Measurement Conference (IMC)*, 2018.

[30] P. Gill, M. Arlitt, Z. Li, and A. Mahanti. The flattening internet topology: Natural evolution, unsightly barnacles or contrived collapse? In *International Conference on Passive and Active Network Measurement*, pages 1–10, 2008.

[31] V. Giotsas, M. Luckie, B. Huffaker, and K. Claffy. Ipv6 as relationships, cliques, and congruence. In *International Conference on Passive and Active Network Measurement*, pages 111–122. Springer, 2015.

[32] F. Gont and T. Chown. Network Reconnaissance in IPv6 Networks. RFC 7707 (Informational), Mar. 2016.

[33] P. Hick and J. Polterock. Ipv6 adoption as seen from an internet backbone link, 2018. `https://blog.caida.org/best_available_data/2018/05/29/ipv6-adoption-as-seen-from-an-internet-backbone-link/`.

[34] R. Hinden and S. Deering. IP Version 6 Addressing Architecture. RFC 4291 (Draft Standard), Feb. 2006.

[35] B. Huffaker, M. Fomenkov, and k. claffy. Internet topology data comparison. 2012.

[36] G. Huston. Bgp routing table analysis, 2018. `https://bgp.potaroo.net/index-v6.html`.

[37] Y. Hyun and k. claffy. Archipelago measurement infrastructure, 2018. `http://www.caida.org/projects/ark/`.

[38] k. claffy, Y. Hyun, K. Keys, and M. Fomenkov. Internet mapping: from art to science. In *IEEE Cybersecurity Applications for Homeland Security*, Mar. 2009.

[39] E. Katz-Bassett, J. P. John, A. Krishnamurthy, D. Wetherall, T. Anderson, and Y. Chawathe. Towards ip geolocation using delay and topology measurements. In *Proceedings of the 6th ACM SIGCOMM conference on Internet measurement*, pages 71–84, 2006.

[40] K. Keys. Internet-scale ip alias resolution techniques. *SIGCOMM Comput. Commun. Rev.*, 40(1), 2010.

[41] J. Lee. Using guided missiles in drive-bys. DEFCON-17, 2009. `https://www.defcon.org/images/defcon-17/dc-17-presentations/defcon-17-egypt-guided_missiles_metasploit.pdf`.

[42] M. Luckie. Scamper: a scalable and extensible packet prober for active measurement of the Internet. In *IMC*, Nov. 2010.

[43] M. Luckie, R. Beverly, W. Brinkmeyer, and kc claffy. Speedtrap: Internet-Scale IPv6 Alias Resolution. In *Proceedings of the ACM SIGCOMM Internet Measurement Conference (IMC)*, pages 119–126, 2013.

[44] M. Luckie, A. Dhamdhere, D. Clark, B. Huffaker, et al. Challenges in inferring internet interdomain congestion. In *Proceedings of the 2014 Conference on Internet Measurement Conference*, pages 15–22, 2014.

[45] M. Luckie, B. Huffaker, A. Dhamdhere, V. Giotsas, et al. As relationships, customer cones, and validation. In *Proceedings of the 2013 conference on Internet measurement conference*, pages 243–256, 2013.

[46] J. Martin, D. Rhame, R. Beverly, and J. McEachen. Correlating GSM and 802.11 Hardware Identifiers. In *Military Communications Conference (MILCOM)*, pages 1398–1403, Nov. 2013.

[47] J. Martin, E. C. Rye, and R. Beverly. Decomposition of MAC Address Structure for Granular Device Inference . In *Proceedings of the 32nd Annual Computer Security Applications Conference (ACSAC)*, Dec. 2016.

[48] A. Murdock, F. Li, P. Bramsen, Z. Durumeric, and V. Paxson. Target generation for internet-wide ipv6 scanning. In *Proceedings of ACM IMC*, 2017.

[49] T. Narten, R. Draves, and S. Krishnan. Privacy Extensions for Stateless Address Autoconfiguration in IPv6. RFC 4941 (Draft Standard), Sept. 2007.

[50] R. Padmanabhan, A. Dhamdhere, E. Aben, N. Spring, et al. Reasons dynamic addresses change. In *Proceedings of the 2016 Internet Measurement Conference*, pages 183–198. ACM, 2016.

[51] R. Padmanabhan, Z. Li, D. Levin, and N. Spring. Uav6: Alias resolution in ipv6 using unused addresses. In *International Conference on Passive and Active Network Measurement*, pages 136–148, 2015.

[52] D. Plonka and A. Berger. Temporal and spatial classification of active ipv6 addresses. In *Proceedings of ACM IMC*, 2015.

[53] D. Plonka and A. W. Berger. kIP: a Measured Approach to IPv6 Address Anonymization. *CoRR*, abs/1707.03900, 2017.

[54] RIPE NCC. RIPE Atlas, 2018. `https://atlas.ripe.net/`.

[55] RIPE NCC. RIPE Atlas v6 Measurements, 2018. `https://atlas.ripe.net/api/v2/measurements/6152/`.

[56] J. P. Rohrer, B. LaFever, and R. Beverly. Empirical Study of Router IPv6 Interface Address Distributions. *IEEE Internet Computing*, Aug. 2016.

[57] X. Shi, Y. Xiang, Z. Wang, X. Yin, and J. Wu. Detecting prefix hijackings in the internet with argus. In *Proceedings of the 2012 Internet Measurement Conference*, pages 15–28, 2012.

[58] I. Society. State of ipv6 deployment, 2018. `https://www.internetsociety.org/wp-content/uploads/2018/06/2018-ISOC-Report-IPv6-Deployment.pdf`.

[59] N. Spring, R. Mahajan, and D. Wetherall. Measuring ISP topologies with Rocketfuel. *ACM SIGCOMM Computer Communication Review*, 32(4), 2002.

[60] S. Thomson, T. Narten, and T. Jinmei. IPv6 Stateless Address Autoconfiguration. RFC 4862 (Draft Standard), Sept. 2007. Updated by RFC 7527.

[61] V. Tran, Tony. Ipv6 geolocation using latency constraints. Master's thesis, NPS, 2014.

[62] Y. Vanaubel, P. Mérindol, J.-J. Pansiot, and B. Donnet. A brief history of mpls usage in ipv6. In *International Conference on Passive and Active Network Measurement*, pages 359–370, 2016.

[63] Y. Vanaubel, J.-J. Pansiot, P. Mérindol, and B. Donnet. Network fingerprinting: Ttl-based router signatures. In *Proceedings of the 2013 conference on Internet measurement conference*, pages 369–376, 2013.

[64] Verizon. Ipv6 at verizon wireless, 2017. `https://www.apnic.net/wp-content/uploads/2017/01/vzw_apnic_13462152832-2.pdf`.

[65] Y. Wang, D. Burgener, M. Flores, A. Kuzmanovic, and C. Huang. Towards street-level client-independent ip geolocation. In *NSDI*, volume 11, pages 27–27, 2011.

[66] W. Willinger, D. Alderson, and J. C. Doyle. Mathematics and the Internet: A source of enormous confusion and great potential. *Notices of the AMS*, 56(5), 2009.