ABSTRACT
The IPv4 address space is small enough to allow exhaustive active measurement, permitting important insight into Internet growth, policy, and evolution. The IPv6 address space, on the other hand, presents the problem that we can no longer perform exhaustive measurements in the same way, inhibiting our ability to continue studying Internet growth. Access to private datasets (e.g., HTTP access logs on content servers, flow data in ISP networks, or passive DNS traces) solves some problems but may not be feasible or desirable. This paper describes IPv6 address collection by exhaustively sweeping the reverse DNS domain for the IPv4 address space and performing AAAA queries on the results. Subsequent ICMP and TCP measurements are conducted to measure the responsiveness of the resulting set. Key outcomes include: the PTR sweep discovers 965,304 unique, globally routable IPv6 addresses originating from 5,531 ASNs. 56% of the addresses are responsive, across 4,571 ASNs. Upon inferring pairs of IPv4 and IPv6 addresses that are likely associated with the same device, the data indicates a trend toward IPv4 addresses being more responsive than their IPv6 counterparts, with a higher incidence rate of TCP connections being refused, and wide disparity on where TCP connections or ICMP echo requests fail silently when comparing IPv4 and IPv6. The disparity in IPv4 and IPv6 responsiveness is highly variable, and indicative of distinct host configuration and network policies across the two networks, presenting potential policy or security gaps as the IPv6 network matures.

1. INTRODUCTION
Exhaustive active measurement of the Internet’s IPv6 address space is infeasible: the current global allocation defines up to $2^{125}$ addressable hosts and $2^{61}$ addressable networks [2]. Individual IPv6 BGP advertisements are intractably large for exhaustive scans: BGP tables from February 29th 2016, 43.5% of advertisements were /48s (i.e., $2^{16}$ individual /64s); 25.3% were /32s. The pace of growth in the IPv6 network necessitates additional sources of addresses for direct measurement, or to augment existing heuristics used to constrain the search space for active measurement.

The problem for active measurement in an IPv6 Inter-
2. BACKGROUND

Other work has studied aspects of IPv6 deployment and maturity. Czyz et al. studied various metrics IPv6 deployment metrics [8], including nameservers with IPv6 connectivity, IPv6 glue records, queries arriving at nameservers, and others including active measurement on the top 10,000 domains listed in the Alexa top-million popularity index. The study shows that IPv6 is maturing by all metrics, though not all at the same pace.

Plonka and Berger contributed traffic measurements from Akamai’s content delivery network, and characterised the temporal stability of the IPv6 space in [8]. Their study features IPv6 addresses collected from HTTP access logs, and shows how hosts or ISPs use that space.

On network maturity, Livadaria et al. [16] compare aspects of IPv4 and IPv6 network stability, including routing stability and the effect on data-plane stability. Their findings suggest that the IPv6 network is proportionally less stable than the IPv4 network. Beverly et al. also recently studied router availability, albeit without a contrast to IPv4 stability [6]; much earlier work attempted to perform global topology discovery [20, 15].

Understanding the structure of the IPv6 Internet, and the means by which anybody may be able to conduct network measurement or by which attacks may be staged, is important; for example, how the IPv6 network structure may be used to launch malicious traffic [5]. Other work has listed speculative approaches or concerns around to active IPv6 measurement [12].

Some work has attempted to use the ip6.arpa DNS domain to locate public IPv6 addresses [10], though the approach may not return many responses [14]. Nikkhah et al. also used the Alexa top-million index to feed DNS queries to locate A and AAAA records to perform active measurements on connectivity and throughput [19].

Earlier work suggested that stateless addressing (SLAAC) or addressing that placed a device’s IPv4 address into the bottom 32-bits of its IPv6 were commonplace [17]. Since then, privacy extensions for autonomous host addressing were specified [18] and have become common primarily in end-hosts. RFC 7721 [7] provides an overview of privacy, and address generation mechanisms.

3. SWEEPING IN-ADDR.ARPA

The first step of this work is to collect DNS names from which AAAA records can be queried. The starting point is understanding that, first, the IPv4 space is currently under high utilisation and, second, that performing queries to retrieve PTR records for all IPv4 addresses is feasible and cheap. The in-addr.arpa domain is used as a means to perform a query against an IPv4 address, returning one or more names for a host if the owner of the address space has provided one.

The principle caveat is to note that there is no operational requirement for network administrators to configure PTR records for their address space. While coverage will not be complete, the practice is currently common enough to provide a large set of names. In many cases, a successful query for a PTR record is likely to return a name that refers to a single network device.

3.1 Approach

To conduct all the DNS queries presented in this section and the active measurements presented in Section 5, four virtual hosts hosted by DigitalOcean were deployed in the UK. Each VM runs a local instance of bind, reachable only by processes on the localhost.

Each host ran a program dedicated to performing the DNS queries presented in this study (A, AAAA, and PTR), issuing and handling multiple DNS queries asynchronously with the assistance of libevent.

For each IPv4 address derived from a full routing table collected by Route Views on February 29th 2016 [3], a PTR query was sent and the responses, including error codes, stored for inspection. This table includes 2,814,910,336 global IPv4 addresses, each of which is used to create a reverse DNS query for this study. The AAAA responses are covered in Section 4.

DNS PTR queries took place between March 8th and March 17th 2016. DNSSEC was not used during this initial study, but should be enabled for future work. In cases where the DNS response is too long for a UDP datagram, queries are reissued over TCP.

3.2 PTR Responses Obtained

A summary of responses is outlined in Table 1(a). 1.19 billion queries returned names (around 42.8%). 1.4 billion return an NXDOMAIN error, with no configured subdomain for the address space. Around 200 million returned a server failure code, and another 3 million where the domain is configured but no record is found. A small number of queries failed with timeouts. From the set of 1.4 billion queries that returned an answer, a short breakdown of obvious misconfigurations of bad data is presented in Table 1(b).

24,945 names returned multiple PTR records. A partial distribution of the larger answer sets is shown in Figure 1, indicating that after the common convention of one record per name, the most common PTR sets, though rare, are 15-20 records in size. The largest set observed from a PTR query was 1,248 records associated to one IPv4 address. Breaking out the records that returned multiple responses, we have a full set of 1,190,767,539 names.

The names revealed through this process often identify infrastructure nodes (routers, middleboxes, firewalls, etc) in addition to hosts intended for use as public servers. Common strings surface: “static”, “customer”, “gw”; numeric strings and two or three-character codes. The set of names is therefore distinct from datasets
Figure 1: Partial range of PTR response sets larger than one record; full range includes sets with over 1,000 records.

Table 1: PTR query response overview.

| Type             | Count     | String   | Count |
|------------------|-----------|----------|-------|
| No domain        | 1,421,766,914 | localhost | 2,401,398 |
| Serv failure     | 199,697,905  | empty string | 965,114  |
| No data          | 3,055,458   | IPv4 addr | 184,858 |
| Other            | 32,112      | 0.0.0.0  | 1,517  |
| No Error         | 1,190,362,811 | 0.0.0.0  | 10     |
| Total            | 2,814,915,200 | (b)       |        |

Table 2: Overview of results to the AAAA queries.

| Network Name   | ASN  | #    | Country | Count |
|----------------|------|------|---------|-------|
| SURFnet (NL)   | 1103 | 93,649 | DE      | 237,896 |
| Deutsche Tele. (DE) | 3320 | 79,903 | US      | 222,528 |
| 1&1 (DE)       | 8560 | 51,326 | EU      | 123,423 |
| Comcast (US)   | 7922 | 43,384 | RU      | 43,679  |
| GMO (JP)       | 7506 | 40,672 | GB      | 40,082  |
| Yandex (RU)    | 13238| 40,672 | FR      | 33,733  |
| Host Europe (DE) | 20773| 23,522 | NL      | 28,547  |
| Hetzner Online (DE) | 24940| 19,404 | CZ      | 15,949  |
| Contabo (DE)   | 51167| 15,516 | NO      | 11,355  |
| CloudFlare (US) | 13355| 14,387 | SG      | 11,198  |

Table 3: Networks and countries with the greatest number of unique IPv6 addresses discovered.

| Network Name  | ASN | #    | Country | Count |
|---------------|-----|------|---------|-------|
| unknown.level13.net |      |      |         |       |

4. IPV6 ADDRESS SWEEP

The names discovered in the previous section form the basis for the next stage, querying for AAAA records.

4.1 Approach

Using the same infrastructure as the name collection stage, a AAAA query was issued for each of the 1,190,767,539 names, and the responses were collected. These queries were conducted between March 24th and March 28th. Results from the previous section were not de-duplicated, but bind was configured to cache results to minimise the number of duplicate queries emitted.

4.2 Results

An overview of the responses in this stage is shown in Table 2(a); 4,742,818 queries (0.4%) returned AAAA records, with particular responses shown in Table 2(b): for example, 2.4 million responses were the result of querying the “localhost” string from the PTR sweep, and a further million came from a common string.

Table 4 describes sets of IPv6 addresses discovered that match the blocks defined in the IANA special...
Table 4: Overview of special-purpose addresses.

| Class                        | Count   | Class   | Count |
|------------------------------|---------|---------|-------|
| 6to4                         | 109,078 | ::      | 84    |
| v4-mapped Addr               | 5,080   | Teredo  | 72    |
| ::1, not “localhost”         | 4,350   | Documentation | 18  |
| Link-local                   | 981     | IETF    | 5     |
| IPv4-v6 Translat.            | 555     | Direct, AS112 | 3   |
| Unique-Local                 | 553     |         |       |

Stable IIDs: 73,532 addresses were returned with bits 7 and 25 – 39 (“ff:fe”) set in the IID, indicating a SLAAC address generated from a MAC address. 72,888, are globally routable, associated with hosts from 1,007 different ASNs. RFC 7043 [4] states that these static addresses should not be published in the DNS because of the the privacy concern they present.

Special-Purpose Addresses: Table 4 lists the number of names that resolve to special-purpose IPv6 address ranges [1]. In all, 120,780 queries returned a special-purpose address. Some 304 names returned both globally routable addresses and special-purpose addresses.

Non-standard Addresses: 28,889 unique responses fall outside the standard IANA allocations. Many of those are the origin IPv4 address in the bottom 32-bits without the network portion set, much like the deprecated IPv4-Compatible IPv6 Addresses [13]. Many others are 32-bit values padded into the most significant bits of the returned address, which do not appear to match the origin IPv4 address.

::1: The origins of the localhost strings in Table 2 is largely constrained to a small set of ASNs, primarily registered in Vietnam; the largest contributors are ASNs 45899, 7552, and 7643.

4.3 Pairing IPv4 and IPv6 Addresses

Next, we attempt to consider IPv4 and IPv6 addresses in pairs. Attempting to pair addresses across protocol families offers scope for direct like-for-like measurement of performance or network policy, against infrastructure which is not as heavily monitored as, say, content servers. It may be possible to determine the gaps in host configuration and security policies between the IPv4 and the IPv6 networks.

To pair IPv4 and IPv6 addresses, the following approach was used: in each of the cases where an AAAA query returned at least one result which was a globally routable address with an ASN in the routing table used throughout this paper, an additional A query was attempted on the same domain name. There are of course many cases where names may resolve to multiple addresses (for either family); to pair addresses, I have conservatively retained only IP addresses following the following criteria: addresses resolve from the same name, originate from the same ASN, and where no address resolves against any other name with a contradictory result. Note that this could be expanded in cases where the same organisation uses multiple ASNs.

This process leads to 673,108 pairs of unique IPv4 and IPv6 addresses originating from 5,228 ASNs. Note there is no requirement that A and PTR records transpose: querying the PTR record for a given IP ad-
dressing then the A record of the resulting name need not return the original IPv4 address. Further, the name in the PTR record may have A and AAAA records, but without any requirement that they map onto the same network device. In the common case, however, it is likely that there is a strong correlation of A records and AAAA records mapping onto a single device or interface, not least to reduce management complexity.

4.4 Summary

Although the yield on this form of collection is low, the range of addresses obtained and the location of those (by ASN, registered country, and set of ISPs) is broad. As a data source for forming sets of addresses for measurement, range is important. Broad coverage such as this helps us to understand the subnets actually in use as a subset of what is visible in BGP, and to better understand static address allocation patterns within networks. The addresses found here are used in the following section.

5. ACTIve MEASUREMENT

In this section, we will use a limited set of active measurements to better understand how active the set of addresses is. These measurements are intended as a lightweight first attempt at characterising the addresses, rather than anything that may be construed as invasive port scanning or flooding.

While tools exist specifically for either port scanning (nmap) or rapid measurement of the IPv4 address space (zmap [11] and masscan), tooling for large-scale scanning is weaker for IPv6. scan6 attempts to speculate search IPv6 networks with various heuristics; in this work, we have a fixed set of target addresses.

The two broad tests conducted on these addresses were ICMP echo requests, and TCP connection attempts to various well-known port numbers. These tests used the same virtual machines as used for the DNS queries, employing GNU Parallel with standard tools: ping and ping6 for the echo requests, and the standard OpenBSD release of netcat for TCP connections.

The set of unique IPv6 addresses was reordered using the GNU tool shuf, and the paired IPv4 addresses identified in Section 4.3 were inserted adjacent to their IPv6 counterpart. This list was divided across the four virtual machines, and each individual step scheduled to take approximately 24 hours. The intention here is to be deliberately lightweight, taking measurements against apparent pairs at approximately the same time but otherwise attempting to stage the work such that consecutive runs of addresses are avoided.

5.1 Ethical Considerations

The measurements for this study were all conducted against IP addresses publicly listed in DNS: no brute-force or speculative measurement was attempted. The measurements relied on no unusual TCP or IP options, didn’t send any data aside from packets required for echo requests, TCP handshakes, and TCP tearowns. TCP connections were closed immediately if successful, and no data from the connection was read or stored. Nothing other than the outcome of an echo request or a connection attempt is stored (packets, payload, etc, are all discarded immediately).

The hosts running the measurements also each ran a webserver, configured to serve an informational page on the study with contact details. Five networks requested, via DigitalOcean, that the measurements cease. In these cases, all addresses advertised from the ASN identified in the report from future measurements.

5.2 ICMP Echo Requests

Three ICMP echo requests were sent to each target address, the address marked active if any responses were received. Timeouts of 3s were recorded.

544,156 addresses (56.7%) responded to echo requests, across 4,573 ASNs (82.7% of the ASNs discovered in Section 4). The range of responses by ASN, as a percentage of the set of IP addresses per ASN, is indicated in Figure 3; the clear bimodal pattern is generated by ASNs with small sets of hosts either all responding or not responding at all, while the ASNs with larger address sets offer a much more varied response.

5.3 TCP Measurements

TCP connections were attempted to the addresses discovered in the previous section, on a range of port numbers: 21, 22, 53, 80, 443, and 8080. The purpose is not to be exhaustive, merely to investigate common port numbers to determine if there are gaps in policy or configuration. The full list of addresses was tested against one port before progressing to the next. Connections were staged slowly, paced at around 85,000 addresses an hour. When a TCP connection timed out (using a 3s timer), no retry was attempted.

Broadly, on port 80, around 21.1% of connections were made successfully, and 13.3% were refused; on port 22, 18.5% were successful and 9.4% refused. In all cases, around 60% of requests were silently dropped.

![Figure 3: Histogram showing distribution of ASNs responding to echo requests.](image-url)
Notably, 31,527 addresses accept or reset a TCP connection (on at least one of the ports tested) in cases where ICMP echo requests were not returned. This may be intentional, if network administrators have decided to drop ICMP traffic to or from hosts against the current practice specified in [9].

5.4 Evaluating the IPv4-IPv6 Pairing

Here, we will refer to the pairs of addresses identified in Section 4.3. Figure 4(a) presents an overview of the response rates for echo requests and TCP connection attempts on the port numbers tested. First, note that on this smaller set of addresses, the ICMP echo response rate is slightly higher than the full set of addresses available, with 61.7% of the paired IPv6 addresses responding. The response or success rates on all tests was lower on IPv6 than IPv4, however: 67.8% of the IPv4 addresses responded to echo requests.

Figure 4(b) attempts to break down the distributions of the response rates in each ASN according to the type of the response. In all cases, IPv4 hosts are more likely to respond to ICMP echo requests, and are more likely to accept connections on the ports tested, while the IPv6 devices are more likely to reject connections, implying services are configured to listen on IPv4 but not yet on IPv6. TCP drop ratios are, on average, equivalent for IPv4 and IPv6 traffic, though the spread in individual networks is wide. These trends are not universal, and there is wide disparity in responses across networks. The data presented here suggests there are gaps in host or network policy between IPv4 and IPv6; in some cases this may be intentional, for example if server software isn’t deemed stable with IPv6 traffic.

6. CONCLUSIONS/IMPLICATIONS

This paper has presented an approach for discovering IPv6 addresses that can be used for public measurement. By sweeping PTR records of the IPv4 address space as advertised in BGP then performing AAAA queries against the returned names, 965,304 unique IPv6 addresses are located in 5,531 autonomous systems. 56.7% of those respond to ICMP echo requests from 82.7% of those autonomous systems.

On attempting to pair IPv4 and IPv6 addresses, 673,108 pairs of unique IPv4 and IPv6 addresses originating from 5,228 are discovered in ASNs. 61.7% of the IPv6 addresses respond to echo requests compared to 67.8% of the IPv4 addresses. Performing TCP measurements against these addresses exposes a similar trend: devices are more likely to respond over IPv4 than over IPv6. The causes of this are unclear, but a slightly higher proportion of IPv6 hosts refusing connections implies host configuration for services lagging network policies.

While this approach relies on the IPv4 space to locate IPv6 addresses on which measurements can be attempted, the IPv4 network is a resource that we should not ignore while it is still dominant. Obviously, approaches must change in future when IPv6 becomes the dominant protocol family.

This study highlights some configuration, privacy, or security concerns. For example, SLAAC addresses in the DNS may be intentional or accidental. Similarly, dropped ICMP traffic may be intentional, or indicative of immature network security policies.

Finally, this study has shown some structure evident in the IPv6 addresses collected. Bit patterns from real IPv6 deployments, especially for fixed infrastructure, is useful to help improve existing heuristics for speculative active measurement studies. Such heuristics are by definition not exhaustive, but may allow active measurement studies of the IPv6 space similar to the large body of existing IPv4 active measurement work.

7. ACKNOWLEDGEMENTS

Jonathan Balkind, for reviewing earlier reversions of this paper.
8. REFERENCES

[1] IANA IPv6 Special-Purpose Address Registry. https://www.iana.org/assignments/iana-ipv6-special-registry/.

[2] Internet Protocol Version 6 Address Space. https://www.iana.org/assignments/ipv6-address-space/.

[3] University of Oregon Route Views Project. http://routeviews.org/bgpdata/2016.02/RIBS/rib.20160229.2200.bz2.

[4] J. Abley. Resource Records for EUI-48 and EUI-64 Addresses in the DNS. RFC 7043 (Informational), Oct. 2013.

[5] S. M. Bellovin, B. Cheswick, and A. Keromytis. Worm propagation strategies in an IPv6 Internet. LOGIN: The USENIX Magazine, 31(1), 2006.

[6] R. Beverly, M. Luckie, L. Mosley, and kc claffy. Measuring and Characterizing IPv6 Router Availability. In Proceedings of the 16th International Conference on Passive and Active Measurement, 2015.

[7] A. Cooper, F. Gont, and D. Thaler. Security and Privacy Considerations for IPv6 Address Generation Mechanisms. RFC 7721 (Informational), Mar. 2016.

[8] J. Czyz, M. Allman, J. Zhang, S. Iekel-Johnson, E. Osterweil, and M. Bailey. Measuring IPv6 Adoption. In Proceedings of the Annual ACM SIGCOMM, 2014.

[9] E. Davies and J. Mohacsy. Recommendations for Filtering ICMPv6 Messages in Firewalls. RFC 4890 (Informational), May 2007.

[10] P. V. Dijk. Finding IPv6 hosts by efficiently mapping ip6.arpa. http://7bits.nl/blog/posts/finding-v6-hosts-by-efficiently-mapping-ip6-arpa.

[11] Z. Durumeric, E. Wustrow, and J. A. Halderman. ZMap: Fast Internet-wide Scanning and Its Security Applications. In Usenix Security, 2013.

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

[13] R. Hinden and S. Deering. IP Version 6 Addressing Architecture. RFC 4291 (Draft Standard), Feb. 2006. Updated by RFCs 5952, 6052, 7136, 7346, 7371.

[14] Q. Hu and N. Brownlee. IPv6 Deployment Concerns in the Global Internet Infrastructure. In IRTF & ISOC Workshop on Research and Applications of Internet Measurements (RAIM)). 2015.

[15] X. Kou and Q. Wang. Discovering IPv6 network topology. In IEEE International Symposium on Communications and Information Technology, 2005.

[16] I. Livadariu, A. Elmokashfi, and A. Dhamdhere. Characterizing IPv6 control and data plane stability. In IEEE INFOCOM, Apr 2016.

[17] D. Malone. Observations of IPv6 addresses. In Proceedings of the 9th International Conference on Passive and Active Measurement, 2008.

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

[19] M. Nikkhah, R. Guérin, Y. Lee, and R. Woundy. Assessing IPv6 Through Web Access A Measurement Study and Its Findings. In Proceedings of the 2011 ACM CoNEXT Conference, 2011.

[20] D. G. Waddington, F. Chang, R. Viswanathan, and B. Yao. Topology Discovery for Public IPv6 Networks. SIGCOMM Computer Communication Review, July 2003.