# Understanding IPv6 Internet Background Radiation

Jakub Czyz, Kyle Lady,
Sam G. Miller,
Michael Bailey
University of Michigan
Ann Arbor, MI
{jczyz, kylelady, sgmiller,
mibailey}@umich.edu

Michael Kallitsis
Merit Network, Inc.
Ann Arbor, MI
mgkallit@merit.edu

Manish Karir
Department of Homeland
Security
Sci. and Tech. Directorate
Cyber Security Division
Washington, DC
manish.karir@hq.dhs.gov

## ABSTRACT

We report the results of a study to collect and analyze IPv6 Internet background radiation. This study, the largest of its kind, collects unclaimed traffic on the IPv6 Internet by announcing five large /12 covering prefixes; these cover the majority of allocated IPv6 space on today's Internet. Our analysis characterizes the nature of this traffic across regions, over time, and by the allocation and routing status of the intended destinations, which we show help to identify the causes of this traffic. We compare results to unclaimed traffic in IPv4, and highlight case studies that explain a large fraction of the data or highlight notable properties. We describe how announced covering prefixes differ from traditional network telescopes, and show how this technique can help both network operators and the research community identify additional potential issues and misconfigurations in this critical Internet transition period.

## Categories and Subject Descriptors

C.2.3 [**Computer-Communication Networks**]: Network Operations—*Network management, Network monitoring*

## Keywords

IPv6; Darknet; Routing; Network Pollution; Internet Background Radiation; Measurement

## 1. INTRODUCTION

We have reached a turning point in IPv6 adoption. The entire IPv4 address space has now been allocated from the Internet Assigned Numbers Authority (IANA) to the Regional Internet Registries (RIRs). Two of these already have run out of their pool of addresses and are reduced to a final /8, from which only very small assignments are given to customers. As a result, IPv4 network growth has slowed to an almost linear rate over time from its historically exponential rate, while IPv6 networks have gone from a linear growth rate in the past to an exponential one [40]. Since World IPv6 Day in June 2011, native IPv6 traffic has doubled [34], and IPv6 traffic constitutes approximately 0.2% of peak bits per second in large cross-provider samples [21].

Historically, network telescopes that capture traffic to "dark" (i.e., unreachable) destinations have frequently been used in IPv4 as sensors to observe phenomena that result as side effects of various applications and events on the Internet (e.g., [31], [38]). For instance, network telescopes can be used to detect worm scanning behavior, backscatter from Denial of Service (DoS) attacks, malicious vulnerability scanning, misconfiguration, and even Internet censorship. Observations of such Internet background radiation serve two purposes: to understand the original phenomenon and to aid in raising awareness regarding issues (e.g., bugs, bad practices, or misconfigurations) that potentially could be fixed. As an example of the latter, one of the results of the work of Wustrow *et al.* was the adoption of a policy to quarantine a particularly polluted block of IPv4 address space from being allocated to the public [38].

As networks attempt to enable an increasing number of IPv6 applications and services, it is highly likely that inexperience, system configuration differences, or even software or hardware bugs will result in errors that can lead to Internet background radiation. Observing such undesired traffic can provide valuable insight to help navigate the rocky start of the new technology. Identifying these issues early in the adoption process minimizes the cost of fixing them and can provide a best-practice template for future IPv6 network operators. In addition, by observing this traffic, we can be alerted to the emergence of malicious activity, such as scanning and worms, via the new protocol.

The primary contribution we make is to report on a large sample of IPv6 background radiation. We conduct the broadest IPv6 network-telescope-based longitudinal study ever preformed by concurrently announcing *BGP prefixes that cover the majority of allocated IPv6 space* used for allocations by each of the five (RIRs)—86% of allocated /64 networks outside of 6to4. We announced the prefixes: 2400::/12, 2600::/12, 2800::/12, 2c00::/12, 2a08::/13, and 2a04::/14 for over a three-month period. For a few days, we also announced RIPE's 2a00::/12. This perspective and scale allows us to capture both spatial and temporal features of the network telescope data. We also compare our results to a week of IPv4 network telescope traffic from unallocated portions of a large ≈/8 prefix captured at the same time as one of our IPv6 collection periods.

Second, we show that a *covering prefix* methodology for network telescopes in IPv6 leads to a much larger sample of data—and of qualitatively different data—than that afforded by a traditional network telescope based on only unallocated address space. We find that 95% of the packets we captured were from allocated space, and we show that captured traffic is clustered close to used prefixes. As the IPv6 address space is particularly sparse (e.g., each IPv6 subnetwork is, by default, 64 bits—an entire IPv4 Internet address space, squared), this methodology is critical to network telescope studies.

*IMC* '13 Barcelona, Spain
Copyright 2013 ACM 978-1-4503-1953-9/13/10$15.00.
http://dx.doi.org/10.1145/2504730.2504732.

Our contributions include:

- A characterization of IPv6 background radiation, including:
  - We observe a mere 1Mbps of traffic in our study.
  - We find significant, qualitative differences between IPv4 background radiation and our observations.
  - We uncover little or no evidence of large-scale malicious scanning, and find that most traffic we see is due to misconfiguration.
- An exploration of the type of traffic observable via a covering prefix methodology, which reveals
  - 95% of the traffic we observed would not have been visible using a traditional network telescopes.
  - This traffic consists of packets misdirected due to IPv6 routing instability (36%) and apparent leakage of address space meant for internal use (59%).

## 2. RELATED WORK

Related work in this area can be divided into the following broad categories: early IPv6 network telescope measurements; IPv4 background radiation analysis; and Internet background radiation collection considerations. In the following sections we discuss the key prior research under each of these three umbrellas.

*Early IPv6 Network Telescope Measurements.* The earliest known study to collect IPv6 background radiation advertised an extremely small /48 prefix for a month in 2006 and only collected 12 packets over that time period [14]. More recent studies have announced a single covering /12 prefix [12,20]. Both of these works served as an initial inspiration for our own experiment. In the shorter term study, Deccio *et al.* collected two weeks of data and reported an average of 74 packets/second of pollution traffic [12]. In the longer term study, Huston *et al.* collected data for 115 days and presented analysis of observed traffic [20]. They reported minimal-to-no increase in average pollution traffic rate during that period. That study also presented a detailed analysis of traffic address destinations. Our results confirm a major finding of that work, which was that the vast majority (≈95%) of captured traffic was for *allocated* networks.

The results we report complement these earlier efforts by scaling up the basic approach in order to better understand global and regional trends. In addition, we also adopt a more rigorous approach by quantifying not only the BGP-control-plane reachability of our prefix announcements, but also data-plane-traffic reachability. Further, we attempted to verify that our global-scale experiment did not negatively impact actual IPv6 traffic on the Internet. Lastly, we present a more thorough analysis of the collected data, providing greater confidence about the generalizability of our results to overall IPv6 pollution.

*IPv4 Background Radiation Analysis.* Wustrow *et al.* analyzed a five-year sample of data collected passively from an unallocated IPv4 /8 block in addition to several newly-allocated-but-unused /8 network telescope prefixes [38]. Their temporal and spatial analysis revealed that background pollution traffic increased four-fold over the course of their sample and that pollution traffic was increasingly dominated by traffic resulting from misconfiguration and other errors.

In work by Pang *et al.*, Yegneswaran *et al.*, and Cooke *et al.*, the authors utilize active responders in order to gain additional insight into the sources of pollution traffic [9,31,39]. Bailey *et al.* discuss the advantages of an active-responder-based network telescope monitor [4]. We did not use active responders in our study in an effort to ensure that we did not directly affect valid IPv6 traffic during this critical transition period; given that we were advertising a *covering prefix* such active response could have interfered with legitimate host traffic.

Glatz *et al.* and Brownlee *et al.* examine sources of Internet pollution and attempt to create classification schemes for this traffic based on various parameters, such as source address, protocol, and inter-arrival times. They present data that quantifies the amount of such traffic on a given network, as well as its properties [7,15].

Internet pollution traffic has also been analyzed to provide insight into large-scale scanning activities [10], Internet censorship [25], and even large-scale physical events such as outages from earthquakes and hurricanes [1, 11]. Barford *et al.* use similar data to find the location of malicious hosts [6].

*Background Radiation Collection Considerations.* Similar techniques to the ones used in our study have previously been discussed in the work of Bailey *et al.*, which describes the use of honeypots with network telescope monitoring [5]. They analyzed a distributed network telescope with over 60 smaller network telescopes and 17 million routable addresses to determine the difficulties in implementing such a hybrid monitoring system.

The size and spatial location of a network telescope are also an important factor in determining the amount of pollution traffic it receives. Moore *et al.* describe the effect the size of a network telescope has on the types of security events witnessed, such as worm spreading, scanning, and distributed denial of service (DDoS) attacks. Similarly, Cooke *et al.* examine Internet background radiation data over ten distributed monitors and study how location effects the collected data [8,30]. These various studies all conclude that location, visibility, route announcement propagation, and filtering all have the potential to effect the observed traffic. Based on these conclusions, we paid particular attention to these data-collection details when designing our study.

## 3. METHODOLOGY

In this section, we describe the design of our experiment and our data collection methodology, as well as the mitigating steps and proactive measurements we conducted to ensure a minimal impact of our covering routes. A long-running study by Huston *et al.* demonstrates that it is possible to conduct a safe IPv6 covering prefix experiment [20]. We sought both to replicate and to greatly expand the scope of that experiment with our work.

### 3.1 The BGP Announcements

In early October 2012, we contacted each of the five RIRs to request permission to announce the entire /12 IPv6 address block that had been allocated to them by IANA. After deliberation, each RIR granted us a Letter of Authority (LOA) temporarily allowing us to announce these prefixes via BGP for the duration of our four-month experiment.

Next, we coordinated with AT&T and Hurricane Electric, the upstream IPv6 providers to Merit Network, Inc., our ISP and research partner. This was necessary to ensure that they would accept our announcements of these unusually large blocks. As anticipated, they both needed to execute special configuration changes (i.e. removing sanity filters) to allow us to make such short covering prefix announcements.

After a series of test announcements, on November 8, 2012 we began announcing all five of the covering /12 prefixes for our experiment. We set up a collection server at Merit Network and advertised it as the final destination for each route. All IPv6 traffic on the Internet that was destined for RIR address space—*but not explicitly claimed by another network via a more specific route*—was routed to our collector and archived. The only large, in-use blocks of IPv6 addresses that fall outside the covering prefixes we advertised were 2001::/12 (used by all RIRs in older allocations), 2002::/16 (used for 6to4 transition), and 2003::/18.

On November 13, RIPE NCC asked us to limit our announcements by withdrawing the /12 route and replacing it with a /13 and a /14, the portion of their address space that hasn't been allocated to customers yet. As explored more fully in the following sections, this change gave us a unique control that helped us discern the nature of Internet background radiation in IPv6. During the course of our experiment, we received four inquiries regarding our BGP announcements from the Internet operations community, which were addressed by providing pointers to a detailed study description. With the exception of the RIPE region's change, we did not need to modify our experiment in throughout the four-month experiment.

## 3.2 Routing Visibility

Since our own providers needed policy changes to accept our routes, we believed that policy could play a role in how broadly visible these routes were. To determine the extent to which the BGP announcements were being accepted and propagated across the broader Internet, we analyzed data from both the Route Views [35] and RIPE NCC Routing Information Service [33] BGP archives.

As seen in Table 1, our announcements were visible from 8 of the 9 IPv6-capable monitors from the Route Views project, including Australia, Brazil, California, Georgia (USA), Japan, South Africa, Virginia, and the UK. The only Route Views monitor that did not see our routes was KIXP in Kenya. In addition, 9 of the 12 IPv6-capable monitors maintained by RIPE NCC saw our announcement, including Austria, California, Italy, Japan, the Netherlands, New York, Sweden, Switzerland, and the UK. We were partially visible by DE-CIX (Germany), which saw two of the six routes (2600::/12 and 2400::/12). Our routes were not visible from MSK-IX in Russia or PTTMetro-SIP in Brazil.

Examining the Route Views peer perspectives more closely, we found that, on average, of the 93 peers at all sites during period A, 74 peers saw our /12 announcements. Likewise, during period B, 75 of the 98 peers saw our /12 announcements. This compares favorably with the number of peers that saw the average IPv6 prefix known to Route Views (66 and 68, respectively). The smaller /13 and /14 RIPE prefixes during period B were just as highly visible for the first half of the period (through mid-January), but only visible to 5 (of 98) Route Views peers in the second half.

Based on our analysis, we conclude that these announcements were visible at the vast majority of IPv6-capable route monitors.

## 3.3 Validating Data Plane Effects

In general, Internet traffic is routed to the most specific prefix in the BGP routing table; therefore, as our experiment consists of announcing shorter (i.e., less specific) prefixes, we would expect to only capture unclaimed IPv6 traffic. However, due to the immature nature of the IPv6 Internet, we were concerned that the longest-match rule, a core routing principle, might not be implemented correctly at every IPv6 node. To identify any potential negative effects, we performed a series of short test announcements (as detailed above) before announcing all of the prefixes. We also closely monitored data plane connectivity during these periods and into the first month of the long-term announcement, as discussed next.

**Table 1: Visibility in Route Views and RIPE Monitors**

| Route Server | LACNIC 2800/12 | ARIN 2600/12 | APNIC 2400/12 | RIPE 2a04/14+ 2a08/13 | AFRINIC 2c00/12 |
|---|---|---|---|---|---|
| **Route Views** | | | | | |
| r-v.eqix | ✓ | ✓ | ✓ | ✓✓ | ✓ |
| r-v.isc | ✓ | ✓ | ✓ | ✓✓ | ✓ |
| r-v.jinx | ✓ | ✓ | ✓ | ✓✓ | ✓ |
| r-v.linx | ✓ | ✓ | ✓ | ✓✓ | ✓ |
| r-v.kixp | | | | | |
| r-v.saopaulo | ✓ | ✓ | ✓ | ✓✓ | ✓ |
| r-v.sydney | ✓ | ✓ | ✓ | ✓✓ | ✓ |
| r-v.telxatl | ✓ | ✓ | ✓ | ✓✓ | ✓ |
| r-v.wide | ✓ | ✓ | ✓ | ✓✓ | ✓ |
| **RIPE RIS** | | | | | |
| rrc00 | ✓ | ✓ | ✓ | ✓✓ | ✓ |
| rrc01 | ✓ | ✓ | ✓ | ✓✓ | ✓ |
| rrc03 | ✓ | ✓ | ✓ | ✓✓ | ✓ |
| rrc04 | ✓ | ✓ | ✓ | ✓✓ | ✓ |
| rrc05 | ✓ | ✓ | ✓ | ✓✓ | ✓ |
| rrc06 | | | | | |
| rrc07 | ✓ | ✓ | ✓ | ✓✓ | ✓ |
| rrc10 | ✓ | ✓ | ✓ | ✓✓ | ✓ |
| rrc11 | | | | | |
| rrc12 | | ✓ | ✓ | | |
| rrc13 | | | | | |
| rrc14 | ✓ | ✓ | ✓ | ✓✓ | ✓ |
| rrc15 | | | | | |

**Table 2: Distribution of the 12,418 IPs used to assess impact of route announcements during testing.**

| RIR Region | No. of IPs | No. of Unique ASNs |
|---|---|---|
| APNIC | 1622 | 603 |
| ARIN | 1219 | 530 |
| LACNIC | 159 | 62 |
| RIPE | 9,409 | 3,654 |
| AFRINIC | 9 | 8 |
| Total | 12,418 | 4,857 |

In order to validate minimal impact on data plane connectivity, we performed the following: We collected a set of public IPv6 addresses by querying the Alexa top 1M domains [2] for AAAA records. We then categorized these by AS number and covering prefix (i.e., RIR region). We conducted ping tests to the resulting set of twelve thousand hosts, which were spread among diverse ASes and regions, summarized in Table 2. *The response rates we saw before and after our announcements were comparable.*

*Port Filtering.* In another approach to validating the representativeness of the captured data, we attempted to detect port blocking between our collection system and sample Internet locations. To this end, we obtained access to a small set of globally-distributed native IPv6 hosts. These hosts were located in Atlanta (AS3595), Japan (AS2516), South Africa (AS33764), Tanzania (AS37084), and the United Kingdom (AS15830). From these hosts, we were able to actively probe our network telescope and examine the data coming into the collection infrastructure. It should be noted that the server in the United Kingdom was unable to have packets routed to the our blocks. However, this appears to be a filtering policy of only the upstream ISP (TelecityGroup Limited), as all of our routes were observed at the London Internet Exchange (LINX, RIS server `rrc01` in Table 1), and AS15830 is a member of LINX.

We began a series of probes of the network telescope prefixes from the remaining four servers. We scanned arbitrarily chosen addresses within the otherwise-unrouted portions of our announcements, such that the scanning packets would be routed to us. We then separated these synthetic packets from organic traffic received by the collector, and discard the scan packets we generated for the rest of the analyses presented here. Upon aggregating the probes, we found *no port-based packet filtering in TCP or UDP* between our four hosts and our collector at Merit Network (AS237). This lies in stark contrast to the experiments conducted by Kreibich, *et al.*, which illustrated widespread port blocking by ISPs for IPv4, as high as 50% for the most commonly blocked ports in IPv4: NetBIOS (TCP/139), SMB (TCP/445), and RPC (TCP/135) [26].

## 3.4 Complications of a Covering Prefix

One interesting difference that emerges in the study of network telescopes that are based on a covering prefix versus traditional ones that are based on unused address space is the impact of routing instability. In a network telescope based on completely *unused* address space, routing instability has no impact, as there are no prefix allocations within the telescope. In the case of a covering prefix, however, the network telescope is effectively the union of all address space (under that prefix) not being otherwise advertised by BGP at any given point in time—whether it is allocated to networks or not. This results in a much more complete telescope but also in complications when the allocated address blocks exhibit instability. In such scenarios, address space can rapidly shift between the network telescope and an allocated and announced address block. As a result, traffic captured due to such instability is much more likely to be composed of otherwise *normal network packets* that happen to be caught by the network telescope due to a (perhaps brief) drop of the more specific route. As such, care must be taken in interpreting and comparing results from network telescopes that do versus ones that do not use a covering prefix announcement. We discuss how we categorized our data with this in mind in Section 3.5.

*Benefits of a Covering Prefix.* During the current transition towards greater adoption of IPv6-enabled networks and services, it is critical to understand any routing instability, as it can help to identify and address the misconfigurations, bad practices, or bugs in software or hardware that are the root cause.

In addition to the ability to observe instability, the main advantage of a covering-prefix-based network telescope study is *much better visibility* due to the known clustering of Internet background radiation near active network prefixes. For instance, Bailey, *et al.* and Cooke, *et al.* have shown that, with an IPv4 network telescope, much more data is gathered when it is located near live hosts [4, 8]. Likewise, Harrop *et al.* discuss, in the context of enterprise IPv4 networks, the advantages of *greynets*, which have unused addresses interspersed with live addresses to produce visibility similar to that of a large, contiguous network telescope [17]. Since the IPv6 address space is vast, compared to IPv4, (and, consequently, sparse) this locality advantage is even more necessary to capture a network telescope sample of any meaningful size. Our results bear this out.

As mentioned above, we initially announced each of the five /12 prefixes that have been assigned by IANA to the RIRs. However, after several days RIPE requested that we reduce our 2a00::/12 announcement to 2a04::/14 and 2a08::/13. Although our reduced RIPE announcement still covered 75% of the initial RIPE address space under the /12, the volume of traffic decreased disproportionately from around 300–900 kilobits per second to 0–80 bits per second for RIPE. On most days, no packets arrived for these RIPE prefixes and we collected an aggregate of only 2,635 RIPE packets over the entire course of the three-month period they were announced[1]. These results agree with previous work in IPv4 [8] and in IPv6 [14, 20], showing more packets near used space.

## 3.5 Data Categorization

We sought to segregate the packets we captured according to whether they would be seen by a traditional network telescope methodology or only by a covering prefix, as we suspected differences. This way, we could characterize traditional Internet background radiation traffic separately from what is otherwise potentially valid traffic to instantaneously unreachable destinations that are normally routed. Likewise, focusing on the traditionally collected traffic would allow an apples-to-apples comparison of this IPv6 data with other IPv4 background radiation studies.

We started by aggregating all of the routed BGP prefixes seen by all Route Views monitors. We combined initial RIBs on the first hour of each experiment period with every subsequent update file (a digest of BGP packets) from all Route Views monitors for each dataset. Since we aimed for a *conservative filtering to produce what we considered traditional background radiation data*, we recorded all prefixes ever announced in any BGP message throughout the entire data collection period, no matter how briefly or sparsely. We excluded from consideration prefixes *shorter* than our own announcements (e.g., a sparse announcement of 2000::/3), as those would not affect our packets.

Next, we aggregated the five RIRs' allocation data and built a list of all prefixes which were allocated by the RIRs prior to the end of each of our two data collection periods. The resulting list in each period, along with the routed prefix list discussed above, constituted our *"allocated"* and *"routed"* filters for that period, respectively. The Cartesian product of these two binary filters gives us four categories of packets.

The type of packets that network telescopes have traditionally captured are the "unallocated and unrouted" category. These are packets destined to prefixes that have not been assigned to any organization by an RIR and which are not normally advertised in the global BGP table—not until the operators of the network telescope announce them. Packets in this category are what have typically been studied when examining background radiation in IPv4 (e.g., the work of Wustrow, *et al.* [38]). Thus, these packets serve as the core data that we use in our characterization of background radiation in IPv6, as discussed in section 4. For succinctness, we term this traffic "dark" in the sections that follow, but we use this term interchangeably with "background radiation". We discuss the other three categories of packets in detail in section 5.

**Table 5: Breakdown of address space under our covering prefixes at the end of February 2013, showing percentage allocated and routed. Also shown is size (in /24 subnets) of non-covering IPv4 network telescope prefix we compare some results to. Note that the size of the entire IPv4 address space is $2^{32}$, which the square root of the size of a single IPv6 /64 subnet.**

| RIR | Prefix | Size of Space in /64 Subnets | Alloc. | Routed |
|---|---|---|---|---|
| APNIC | 2400::/12 | $2^{52} = 4,503,599,627,370,496$ | 3.29% | 1.31% |
| ARIN | 2600::/12 | $2^{52} = 4,503,599,627,370,496$ | 1.85% | 0.20% |
| LACNIC | 2800::/12 | $2^{52} = 4,503,599,627,370,496$ | 6.75% | 0.46% |
| RIPE NCC | 2a00::/12 | $2^{52} = 4,503,599,627,370,496$ | 2.66% | 2.15% |
| RIPE NCC | 2a08::/13 | $2^{51} = 2,251,799,813,685,248$ | 0.00% | 0.00% |
| RIPE NCC | 2a04::/14 | $2^{50} = 1,125,899,906,842,624$ | 0.25% | 0.04% |
| AFRINIC | 2c00::/12 | $2^{52} = 4,503,599,627,370,496$ | 0.43% | 0.41% |
| IPv4 | 35≈/8 | A total of 54,784 /24 subnets | None | None |

## 3.6 Dataset Description

Table 3 summarizes the complete datasets that we captured. We present data from two periods: dataset A (a 24-hour-long capture

[1] While the traffic dropped by three orders of magnitude right away, as mentioned in section 3.2, the two smaller prefixes did have reduced global visibility during the second half of the three-month period as well, which further reduced collected data.

Table 3: Packet counts, rates, and protocol breakdown in the complete datasets, by RIR.

| RIR | Dataset A.Complete (24 hr.; with RIPE /12)<br>2012-11-12T17:00:00 to 2012-11-13T16:59:59 | | | | | | | Dataset B.Complete (3 mon.; with RIPE /13 & /14 only)<br>2012-12-01T00:00:00 to 2013-02-28T23:59:59 | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | kPackets | Average (peak) | | % Protocol Type | | | | kPackets | Average (peak) | | % Protocol Type | | | |
| | | pkts/sec | kbits/sec | UDP | TCP | ICMP | Other | | pkts/sec | kbits/sec | UDP | TCP | ICMP | Other |
| APNIC | 11,826 | 85 (149) | 310 (552) | 48.5 | 4.8 | 46.1 | 0.6 | 1,345,425 | 172 (811) | 523 (1868) | 45.5 | 20.9 | 33.2 | 0.4 |
| ARIN | 12,146 | 88 (517) | 299 (1169) | 72.7 | 3.9 | 23.0 | 0.4 | 2,481,476 | 318 (21185) | 418 (28685) | 23.3 | 67.1 | 9.1 | 0.5 |
| LACNIC | 8,936 | 69 (130) | 210 (370) | 50.4 | 2.1 | 46.4 | 1.1 | 504,848 | 65 (875) | 238 (768) | 28.8 | 4.6 | 65.6 | 1.0 |
| RIPE | 25,490 | 175 (8750) | 235 (6211) | 20.6 | 41.0 | 38.4 | < 0.1 | 3 | < 0.1 | < 0.1 | 8.3 | 27.4 | 64.2 | 0.1 |
| AFRINIC | 298 | 2 (9) | 4 (16) | 56.7 | 7.9 | 35.2 | 0.2 | 20,290 | 3 (88) | 4 (114) | 54.4 | 6.5 | 38.8 | 0.3 |
| Overall | 58,696 | 419 pps | 1.06 Mbps | 41.7 | 19.9 | 38.0 | 0.4 | 4,352,040 | 558 pps | 1.18 Mbps | 30.9 | 45.4 | 23.3 | 0.4 |

Table 4: Categorization of packets by destination in each dataset. "Allocated" are packets with destinations matching a prefix that was allocated by an RIR; "routed" match a BGP prefix known to Route Views any time during collection. Unique number of destinations as well as total TCP payload in each category is also shown.

| Category | Dataset A | | | | Dataset B | | | |
|---|---|---|---|---|---|---|---|---|
| | Packets | % | Unique Dest. | TCP Bytes | Packets | % | Unique Dest. | TCP Bytes |
| **Unallocated, Unrouted ("dark")** | 2,997,540 | 5.11 | 36,855 | 21,880,029 | 208,988,570 | 4.80 | 1,274,798 | 2,901,567,271 |
| **Unallocated, Routed (UR)** | 456 | 0.001 | 28 | 0 | 13,948 | < 0.01 | 848 | 39,216 |
| **Allocated, Unrouted (AU)** | 35,426,005 | 60.36 | 4,784 | 239,960 | 2,576,456,636 | 59.20 | 85,956 | 18,857,335 |
| **Allocated, Routed (AR)** | 20,271,633 | 34.54 | 22,335 | 36,435,501 | 1,566,581,006 | 36.00 | 1,580,052 | 424,971,873 |
| **Total (Complete)** | 58,695,634 | | 64,002 | 58,555,490 | 4,352,040,160 | | 2,941,654 | 3,345,435,695 |

starting on 12 November 2012), and dataset B (a three-month-long capture starting on 1 December 2012). An outage caused by undetected power failure occurred between 5-9 January 2013; we have no packets from that period.

The covering prefixes we announced for each RIR subsumed varying amounts of address space. Table 5 shows the percentage of space under each announced covering prefix that was allocated at any time prior to the last day of the three-month (B) period and that was routed, even briefly, at any time during the period.

*Packet and Dataset Categories.* Table 4 shows the breakdown of the two datasets according to the categories explained in section 3.5. For each category, we also include the number of unique destinations, which gives a sense of the spatial nature of each target address set, as well as TCP payload bytes. The "unallocated and unrouted" subset is the background radiation data that traditional (non-covering prefix) network telescopes have captured.

As Table 4 also shows, due to the covering prefix nature of our route advertisements, we were able to capture a broader spectrum of invalid traffic than would be possible via the traditional network telescope approach. In fact, 95% of the packets we captured were due to our use of a covering prefix. That traffic falls into three categories: unallocated/routed, allocated/unrouted, and allocated/routed. As this traffic is distinct from Internet background radiation as previously studied, we treat it separately and aim to explain and characterize each category in section 5.

*Basic Statistics.* Before we move to a deep analysis of the "dark" subset of our data, we first provide basic high-level statistics about the overall (unfiltered) captured packets. Table 3 shows the volume of all packets per RIR collected during the longer (three-month-long) dataset B time frame (i.e., B.Complete).

An analysis of packet TTL values reveals that the vast majority of received packets (90% captured during time period A and 97% during B) appear to be sent by Windows operating systems (TTL between 64 and 128).

During the 24-hour period selected to be dataset A we received 18.7 GiB in 59M packets. The APNIC, ARIN, LACNIC, and RIPE NCC blocks all received a similar order of magnitude of data, while AFRINIC received two orders less. A similar distribution was observed over the three months that constitute dataset B, except for the loss of nearly all data from RIPE due to withdrawal of 2a00::/12 in favor of two smaller RIPE prefixes. We collected 1,141 GiB via 4.4Bn packets in dataset B. As shown in Table 3, we received an aggregate rate of nearly 1.1 Mbps of traffic during period A and nearly 1.2 Mbps during period B. Since RIPE data in B was nearly zero, we refrain from separately analyzing RIPE in dataset B in the sections that follow, except where noted.

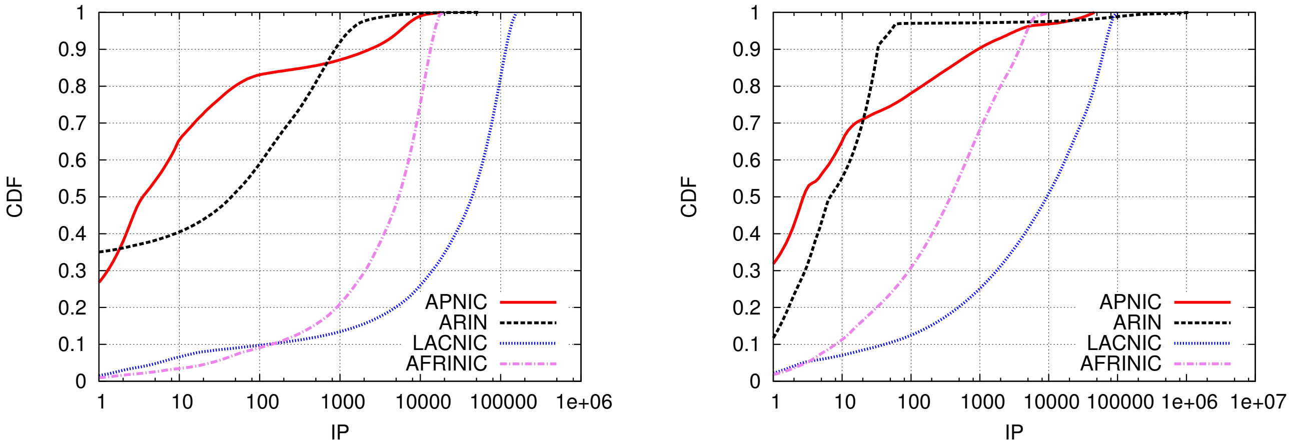
*Protocols.* As Table 3 also shows, the protocol distribution in the overall data is dominated by no single protocol but is made up of between 20 and 50 percent of each of TCP, UDP, and ICMP, though TCP (with 45% of packets) ranks higher than the other protocols in the complete three-month dataset while UDP is the most common protocol in in the unfiltered 24-hour dataset (42%).

*Transition Addresses.* One question of importance when studying the state of IPv6 deployment is the relative proportion of native IPv6 hosts versus those using IPv6 transition technologies. We took advantage of the large and global nature of our collection to explore the distribution of source addresses that were used for 6to4 and Teredo, two of the most prevalent IPv6 transition technologies. We found that, of the 12.5M unique sources observed in dataset B, 12.4% were using 6to4 prefixes. Teredo was used by another 21.7%. Thus, over 34% of sources seen were transition addresses. Note that, since 53% of sources fell under a single native /36 prefix, removing that outlier would yield 72% transition sources. This is approaching the higher 96% seen in an ad-based experiment conducted by Karir *et al.* [23]. We caution, however, that since our sample is largely based on traffic intended for networks that are misconfigured or unstable, it may not accurately reflect the overall IPv6 client population.

# 4. BACKGROUND RADIATION RESULTS

## 4.1 Background Radiation Data Description

As our first aim is to characterize IPv6 background radiation traffic and compare it to IPv4 background radiation, the core statistics we report in this section focus on the unallocated/unrouted ("dark") category of data. We leave explorations of the other three categories of collected packets to the sections that follow.



**(a)**

**(b)**

**Figure 1: Cumulative distribution of source (1a) and destination (1b) IP addresses in B.Dark, sorted by number of collected packets**

## 4.2 Spatial Analysis

In this section, we provide a comparative analysis of spatial features of the dark traffic we observed.

*Traffic Volumes and Patterns.* For the three-month-long background radiation dataset, B.Dark, we observed a total of just 209m packets across all five prefixes. ARIN's traffic dominated the packet count with about 205m packets. We collected about 1m packets in the APNIC dataset. The LACNIC dataset contains 3m packets, and both the AFRINIC and RIPE dataset contain a relatively low number of packets (325k and 2.5k respectively). We refrain from including RIPE in several of the following analyses, due to the negligible number of packets captured after reducing the announcement, as discussed in Section 3.4.

*Source Distributions.* Figure 1a shows the cumulative distribution of source addresses sorted by packet contribution in B.Dark. In both the ARIN and APNIC datasets, we see a few source addresses account for a significant portion of the traffic. In particular, one source address contributes around 30% of the packets, and 90% of these two regions' traffic is accounted for by 1–2k IPs. The other two regions (AFRINIC and LACNIC) both have higher source address diversity, and it takes more than 10k and 100k unique sources to account for 90% of the packets, respectively.

*Destination Distributions.* Figure 1b shows the distribution of destination addresses in B.Dark. Here, we see a slightly greater concentration of destinations and most prominently so for ARIN, where fewer than 30 destination IPs account for 90% of the traffic. In the other three RIRs' datasets, fewer than 10k destinations make up 95% of the packets seen. LACNIC has the most diversity again, with more than 30k unique destinations needed to account for 90% of packets.

## 4.3 Protocol and Port Analysis

*Protocols.* Figure 2a shows the protocol breakdown in the darknet datasets. Just as observed in the discussion of the overall data in Section 3.6, the protocol volumes in the three-month dataset (B.Dark) are heavily biased toward ICMP. However, in the APNIC dataset, we find that TCP dominates the traffic, at 72%, with another 22% of observed packets being ICMP.

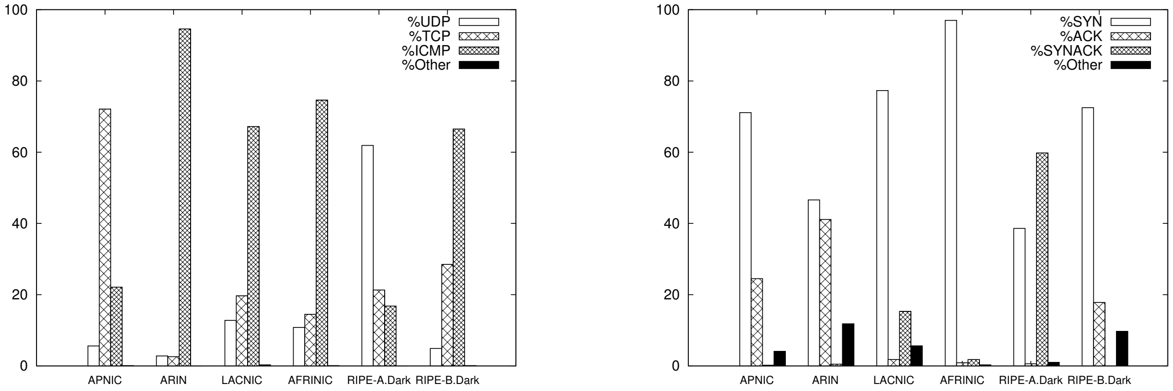
In the one-day dataset with RIPE's network blocks, we see 62% UDP traffic, with TCP and ICMP both contributing about 20%. A surprising 95% of all observed packets in the ARIN dataset were ICMP. We investigate this further in section 4.6.1.

These numbers lie in contrast to typical IPv4 network telescope measurements (e.g., [38]), where TCP dominates, UDP contributes about half as much as TCP, and ICMP is present with only single digits of percentage points of volume. Our own comparison IPv4 analysis, described in Section 4.5, also bears this out, as we see TCP constituting 82% of packets.

*TCP/UDP Ports.* Table 6 presents the top 5 TCP source and destination ports observed in Dataset B.Dark. Source ports 80 and 443 appear in the top 10 or top 20 ports for all datasets. Source port 53 is also the second-most-common port in the ARIN dataset. These suggest that much of the collected traffic is likely to be misdirected responses from DNS and web services. A high number of TCP port 7 packets, used by the Echo protocol to measure round trip times, suggests possible network testing traffic.

Similarly, Table 7 shows the top 5 UDP ports observed in our our different datasets. Port 53, in both source and destination fields, clearly dominates the UDP traffic. While only 6% of UDP packets have source port 53, 28% of the bytes received in UDP packets were in packets sent on port 53. Additionally, nearly 85% of received UDP packets were destined for port 53. These constitute DNS responses and queries, respectively. Another notable pattern in UDP is the significant prevalence of traffic both to and from port 123, which is primarily used for Network Time Protocol packets [29]. Further analysis of some of these features is explored as separate case studies in section 4.6.

*TCP Flags.* The most interesting flag combinations for network telescope research are SYN and SYN+ACK. The former indicates packets where a connection is being attempted (ultimately unsuccessfully to an unreachable address) such as when a host is scanning ports or IPs to connect to. SYN+ACK indicates the response to a previous connection attempt, such as might be seen when an attempt to make a connection is made using a source address that is within the network telescope, and the contacted host replies. When this is done for malicious reasons, it is called backscatter and is often observed during certain classes of DDoS attacks, but this can also be observed in cases of misconfiguration.



(a)

(b)

**Figure 2: Breakdown of Protocols (2a) and TCP Flags (2b) by RIR (from B.Dark unless otherwise noted)**

**Table 6: Top 5 TCP Ports by Packets for Each RIR and Overall in Dataset B.Dark and for RIPE in A.Dark**

| APNIC | | ARIN | | LACNIC | | AFRINIC | | B.Dark Overall, %of TCP Pkts. | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| src | dest | src | dest | src | dest | src | dest | src | % | dst | % |
| 56583 | 80 | 22 | 7 | 445 | 80 | 993 | 39236 | 22 | 30.32 | 7 | 32.40 |
| 49561 | 443 | 53 | 22 | 135 | 45682 | 443 | 45682 | 56583 | 1.59 | 80 | 14.44 |
| 49559 | 2001 | 51211 | 80 | 12829 | 56024 | 143 | 24739 | 445 | 1.22 | 22 | 10.13 |
| 49558 | 445 | 51208 | 34521 | 80 | 61638 | 80 | 26823 | 49561 | 0.59 | 443 | 1.11 |
| 49560 | 5222 | 51207 | 443 | 49155 | 29671 | 5222 | 52232 | 53 | 0.49 | 34571 | 0.85 |

| RIPE (A.Dark) | | | |
|---|---|---|---|
| src | % | dst | % |
| 80 | 35.62 | 179 | 31.10 |
| 443 | 20.45 | 25 | 2.18 |
| 993 | 2.73 | 53 | 1.03 |
| 49166 | 0.85 | 80 | 0.92 |
| 5228 | 0.80 | 40000 | 0.91 |

Figure 2b illustrates the regional differences in the TCP flag combinations in the data we collected. The relatively low volume of SYNACK packets (1.7% in B.Dark) indicates that spoofed-random-source DDoS attacks do not appear to be prevalent in IPv6. However, while low as a percentage of all dark traffic, SYNACKs do constitute a large percentage of the single-day RIPE dataset (59.8%). Our collected TCP packets are overwhelmingly SYN packets; this is most apparent in the AFRINIC dataset, 97% of which are SYN packets. As SYN packets form a plurality, if not the majority, of the TCP packets we see (40.5% in A.Dark and 52.6% in B.Dark), we can infer that a majority of dark TCP traffic coming to the network telescope is connection attempts to unreachable networks.

## 4.4 Temporal Analysis

Overall, we see no clear trend in the volume of IPv6 background radiation over the three-month period of our study. As Figure 3 shows, however, there is a slight decrease in total traffic volume observed by the network telescope in the last two weeks of the B dataset. This is primarily due to a drop in traffic volume to the ARIN region from approximately 30 packets per second to 10. The decreased traffic is related to a reduction in ICMP probing of a small set of ARIN destinations that are heavy recipients of ICMP traffic. We discuss this traffic in section 4.6.1.

## 4.5 Comparison to IPv4

Our work follows a series of seminal IPv4 network telescope experiments that characterized the background radiation of unallocated address space in IPv4, as well as its evolution over time (e.g., [31, 38]). Our IPv6 network telescope results suggest several important differences (and some similarities) compared to that body of work. To produce a more recent and valid comparison, we analyzed a single week of IPv4 background radiation captured during the course of our ongoing IPv6 packet capture.

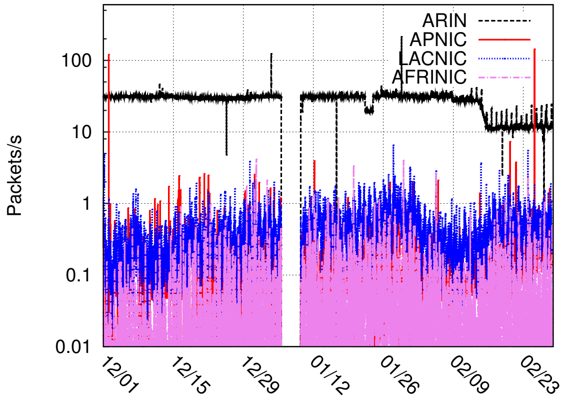


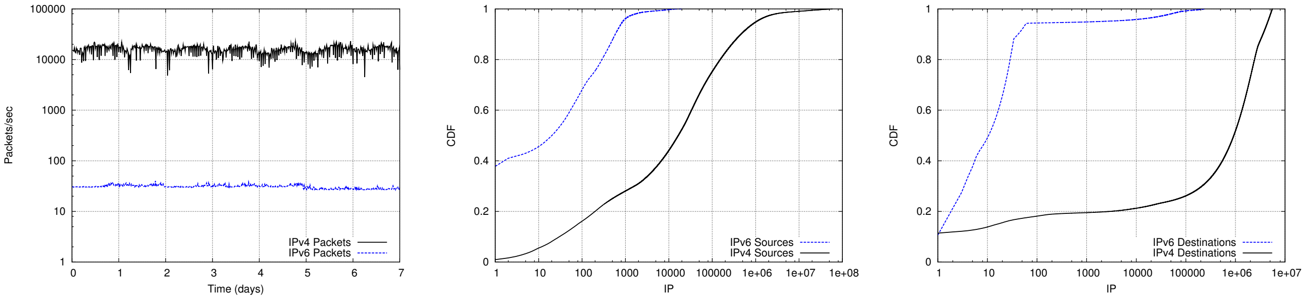
**Figure 3: Background radiation (i.e., dark) packet rate for the four /12s prefixes we announced over three months (dataset B.Dark). The ARIN prefix received about 30 packets per second, whereas the other regions all saw only between 0.1 and 1.0 pps. The hole in January was caused by a power outage.**

The methodology used to capture this data is identical to that described in Section 3.1, with the main differences, aside from the protocol, being that (a) the IPv4 space we monitor is completely *unallocated and unrouted* (by our definition) and thus not a covering prefix; and (b) the size of the address space is considerably smaller: our IPv4 network telescope is composed of address blocks encompassing 13,915,136 total host addresses (≈55k typical subnets). This is equivalent to about 84% of an IPv4 /8, far smaller than our announced IPv6 address space (see Table 5). We subset the three-month background radiation IPv6 data (B.Dark), focusing on the same week (beginning 4 February 2013) for both it and the IPv4 data. We next highlight important differences between the two protocols' background radiation.

**Table 7: Top 5 UDP Ports by Packets for Each RIR and Overall in Dataset B.Dark and for RIPE in A.Dark**

| APNIC | | ARIN | | LACNIC | | AFRINIC | | B.Dark Overall, %of UDP Pkts. | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **src** | **dest** | **src** | **dest** | **src** | **dest** | **src** | **dest** | **src** | **%** | **dst** | **%** |
| 53 | 53 | 53 | 53 | 12929 | 53 | 53 | 39236 | 53 | 5.96 | 53 | 84.81 |
| 16703 | 16703 | 123 | 123 | 53 | 45682 | 45682 | 45682 | 123 | 1.20 | 123 | 1.19 |
| 45682 | 39045 | 33336 | 30718 | 45682 | 3702 | 12829 | 24739 | 12829 | 0.61 | 30718 | 0.71 |
| 12407 | 37385 | 54709 | 33336 | 123 | 123 | 48359 | 26823 | 33336 | 0.31 | 33336 | 0.31 |
| 54593 | 45682 | 500 | 500 | 18600 | 29671 | 20904 | 37648 | 45682 | 0.14 | 45682 | 0.12 |

| RIPE (A.Dark) | | | |
|---|---|---|---|
| **src** | **%** | **dst** | **%** |
| 53 | 10.50 | 53 | 88.60 |
| 32833 | 0.07 | 389 | 0.50 |
| 12589 | 0.05 | 40000 | 0.18 |
| 500 | 0.04 | 16881 | 0.04 |
| 123 | 0.03 | 500 | 0.04 |



**(a) Overall packet volume over the week**

**(b) CDF of source IPs sorted by packets**

**(c) CDF of destination IPs sorted by packets**

**Figure 4: Comparison of IPv4 and IPv6 background radiation during the week starting on 4 February 2013.**

**Table 8: IPv4 and IPv6 protocol breakdown for traffic during the week of February 4, 2013 in the aggregate B.Dark IPv6 dataset and in the IPv4 35.x.y.z network telescope.**

| Darknet Packets | % TCP | % UDP | % ICMP | % Other |
|---|---|---|---|---|
| **IPv4 (35.x.y.z)** | 81.7 | 15.8 | 2.3 | 0.2 |
| **IPv6 (All /12)** | 3.3 | 2.9 | 93.8 | < 0.1 |

*Traffic Volume.* The first striking difference between the two background radiation samples is the relative volume of traffic collected. The IPv6 telescope, as Figure 4a shows, collected an aggregate packet rate of approximately 30 packets per second. In stark contrast, the IPv4 telescope sustained a rate of around 15,000 pps, a 500-fold difference. The bitrate difference (not shown) is similarly unmistakeable. This disparity is expected, due to the still-low adoption of IPv6 and the dearth of malicious traffic over the new protocol. Even the rate of total traffic we received (i.e., with all categories of packets counted), is only about 20-fold higher, at 558 pps. This is in spite of the vastly larger address space covered by our IPv6 announcements as compared to the /8 in IPv4 (as discussed in the previous section).

*Spatial Analysis.* Examining the spatial distributions of the sources and destinations in the two samples, shown in Figures 4b and 4c, respectively, reveals several differences. In the source distributions, we see the IPv4 source contribution ramp-up is slow, taking more than 100 top sources to account for 20% of packets. In the IPv6 data (top line), however, a single source contributes nearly 40% of the packets; we elaborate on that source in Section 4.6.2. The two distributions have similar shape except that the IPv6 line is shifted to the left, as we would expect based on the lower volume of IPv6 traffic, and up, due to the single outlier.

The distributions of the destination IPs, however, suggest a qualitative difference between the two protocols. Namely, the IPv6 sample has many more heavily-hit IPs relative to the total set of destinations, while IPv4 has a small number of heavy hitters. The vast majority of IPv4 destinations only see a relatively low number of packets—e.g., 16.3% of destinations account for just 50% of the IPv4 packets. On the other hand, less than 0.01% (10) of the IPv6 destinations account for 50% of the traffic. We recall that the size of the entire destination address space monitored by the IPv4 network telescope is on the order of $2^{24}$ (nearly 14M unique destinations), whereas the address space of the IPv6 telescope—detailed in Table 5—is on the order of $2^{118}$ hosts, and only a very small percent of routed space generally outside of our visibility. Thus, the fact that so few hosts contacted the IPv6 space we monitored (and that so few IPs are contacted within the space) is itself a significant difference from IPv4.

**Table 9: Top 10 TCP Ports with Percentages in the IPv4 Sample and Comparative Ranking in IPv6 Dark Dataset (if Present) for the Week of 2013-02-04.**

| Source Port | % | v6 Rank | Dest. Port | % | v6 Rank |
|---|---|---|---|---|---|
| 80 | 10.07 | #47 | 445 | 47.75 | #4,268 |
| 6000 | 2.83 | #27,826 | 12350 | 7.14 | |
| 30800 | 0.73 | | 23 | 4.97 | |
| 0 | 0.70 | #5,914 | 80 | 4.15 | #710 |
| 22 | 0.67 | #1 | 443 | 4.03 | #685 |
| 38121 | 0.54 | #20,825 | 3389 | 3.91 | |
| 4935 | 0.51 | | 22 | 2.44 | #442 |
| 7777 | 0.49 | #25,981 | 1433 | 1.68 | |
| 6005 | 0.38 | | 3072 | 1.03 | |
| 21 | 0.36 | | 1024 | 1.03 | |

*Protocol Differences.* In Table 8 we see the protocol breakdown of the IPv4 background radiation compared to that of the aggregated IPv6 B.Dark data for the comparison week. As can be seen, while TCP is the most prevalent protocol in the IPv4 sample (82%), in the IPv6 sample the dominating protocol is, by far, ICMPv6, making up 94% of the packets in the B.dark dataset. This suggests that *there is relatively less pollution traffic in the IPv6 dark space from scanning and backscatter and more from probing, diagnostic, and management traffic.* However, as discussed more in Section 4.6.1, a small block of heavily-hit destinations is responsible for the vast majority of ICMP packets in the dark IPv6 dataset; these inevitably skew these results. That said, as IPv6 is at an early stage of deployment, we certainly expect the properties of its traffic to differ from IPv4.

**Table 10: Top 10 UDP Ports with Percentages in the IPv4 Sample and Comparative Ranking in IPv6 Dark Dataset (if Present) for the Week of 2013-02-04.**

| Source Port | % | v6 Rank | Dest. Port | % | v6 Rank |
|---|---|---|---|---|---|
| 19288 | 1.60 | #51,473 | 10320 | 35.88 | #31,731 |
| 39776 | 1.55 | #18,816 | 5060 | 4.85 | #61,409 |
| 17148 | 1.55 | #39,266 | 47458 | 2.99 | #46,486 |
| 58843 | 1.54 | #12,809 | 53 | 1.99 | #1 |
| 17190 | 1.54 | #43,881 | 137 | 1.39 | |
| 10688 | 1.52 | #48,799 | 3544 | 1.24 | #37,930 |
| 18864 | 1.51 | #1,315 | 39455 | 1.17 | #2,117 |
| 24048 | 1.48 | #25,676 | 65535 | 0.58 | #62,982 |
| 8090 | 1.48 | #60,075 | 1900 | 0.44 | #58,668 |
| 10042 | 1.42 | #59,877 | 161 | 0.35 | |

*Port Distribution.* Tables 9 and 10 show the top ten source and destination ports for each of TCP and UDP, respectively, in IPv4. Next to each port is that port's placement in the corresponding rankings from the IPv6 data during the same week. We do not separately show all the top ports for this week of IPv6 data, as they closely match the top ports seen in the overall dark data (shown in Tables 6 and 7). Looking at the top IPv4 ports, we first observe that, unlike in the IPv6 data where UDP port 53 dominates both directions (78% of destination and 13% of source ports during this comparison week), it is not among the top IPv4 source ports (though it is the fourth destination port). In fact, it ranks as only the 192nd most common source port, accounting for just 0.02% of IPv4 UDP packets. This suggests that relatively little stray DNS traffic is entering the IPv4 network telescope, while this type of traffic is common in our IPv6 data. Indeed, both the prevalence of DNS and ICMP traffic in IPv6 background radiation are congruent with previous findings (e.g., [24]) for overall IPv6 usage, which reported high fractions of these two types of traffic via the new protocol. Other than port 53, we find no coexistent ports in the top 10 lists for UDP in the two samples. For TCP, we also see little similarity, with just port 22 in both protocols' top source port lists and 22 and 80 in both top destination port lists.

Significant absences are TCP/23 (Telnet) and UDP/137 (NetBIOS) from the IPv6 dataset from that week. Despite our earlier discussion of the lack of any port filtering in IPv6 in Section 3.3 and the prevalence of these both in IPv4, we saw zero Telnet packets and only a single NetBIOS packet in that week's IPv6 data. In general, there is a markedly different overall distribution of ports between the two datasets, supporting the hypothesis that *there is a qualitative difference in the nature of traffic in IPv4 and IPv6 background radiation—it is not just a matter of volume*.

*TCP Flags.* Lastly, we compare the proportions of TCP flags in the IPv4 dataset to that in IPv6. As discussed earlier in section 4.3, the most interesting flag combinations for Internet background radiation research are SYN and SYN+ACK. We tabulated SYN and SYN+ACK packets in the IPv4 dataset and found the two categories to make up 83% and 14%, respectively. The IPv6 percentages in this dataset for these two flag combinations are 25% SYN, and 1% SYN+ACK. These differences suggest that *scanning (which manifests itself as SYN packets to dark space) and backscatter (which manifests as SYN+ACK packets to dark space) are both less prevalent in IPv6 background radiation than in IPv4.*

## 4.6 Darknet Case Studies

### 4.6.1 ICMPv6 Probing/Scanning

As described earlier, we see a significant amount of ICMPv6 traffic in our background radiation dataset. Of particular interest to us was whether we would observe signs of large-scale scanning. We see clear evidence of sequential scanning in a handful of cases, though it was generally limited to smaller subnets rather than randomized scans of the entire address space.

Focusing on just the background radiation subset of the data, we find 16 APNIC, 1,646 ARIN, 9 LACNIC, and 3 AFRINIC addresses sourcing over 1,000 ICMP packets in B.Dark. Of these, we find 249 addresses in ARIN with over 100k ICMP packets and five with over a million. Of those five, one begins with fe80:: (discussed in Section 4.6.2), while two are in Akamai Technologies address space.

An interesting specific example is the single ARIN Region IP address 2600:140f:b::b81a:a21f, which is also from address space belonging to Akamai. It generated 2.5M total ICMPv6 packets to 141 unique destinations. Other IP addresses from the Akamai address space sent similarly large amounts of ICMPv6 traffic that was received by our network telescope. In total, we observed over 17M ICMPv6 packets originated by Akamai in B.Dark. We speculate that Akamai probes hosts that provide IPv6 content when making redirection or other content-update decisions.

In one extreme case, we saw fe80::224:38ff:fe7e:af00, a single, link-local ( [18]) IP address, send over 71M ICMPv6 packets to only 27 unique destinations, all within the same /120 prefix (i.e., with only the last 8 destination bits varying) under 2607:fc86::/32. A similar number of packets (around 2.6M) were sent to each address, suggesting a repeated automated scan of this small set of destinations. While we do not know the source of these particular packets, it is possible that the misconfigured host was attempting to monitor the availability of a small group of devices, except that the address range of those devices is in neither publicly advertised space nor allocated space. Interestingly, this is not the only source address probing those same destination addresses.

Upon deeper inspection, we find that hosts within the same /120 subnet along with a closely-addressed second /120, both under the same 2607:fc86::/32 block, were probed by an additional 100M packets. This addition brings the total amount of traffic destined for this unallocated block to 192M packets. Thus, *a staggering 92% of our total background radiation packets was ICMPv6 traffic to or from these hosts*. There are over 3,500 unique source IPs that sent traffic to these two /120 groups, which totaled just 66 destination addresses. The sources originate from a wide swath of 378 different /32s, mostly in ARIN space but with some in each region.

We found that the size of these ICMPv6 payloads was either 10 (about 40% of the packets), 1,000 (30%), or 64 (30%) bytes. Sixty-three percent of the packets are echo requests, and the rest are echo replies. This suggests that traffic from hosts addressed in this block is being sent and with solicitation of replies, which we end up capturing. Indeed, an examination of the other datasets reveals some packets sourced from hosts in this block, which we end up capturing because they themselves happen to contact, for example, allocated-but-unrouted addresses. For example, this is the case for four hosts in this range sending around 5k packets in the B.AU dataset and one host sending ten packets in the B.AR dataset.

While we are unable to track down the location of the hosts using this invalid address block, the case study highlights how a single misconfigured network prefix can cause a large volume of pollution—92% of our background radiation packets.

### 4.6.2 Link-local Leakage

We mentioned above that we saw a very large fraction of packets from the IPv6 address fe80::224:38ff:fe7e:af00. This address's packets were the largest single contributor to our background radiation data, accounting for 34% of all received packets.

As addresses beginning with $fe80$ are "link-local" and meant to only be valid on local networks, their presence in our collection indicates a misconfiguration [18]. This address appears to use the EUI-64 encoded MAC address format for the host portion of the address (indicated by the FF:FE in the middle of the last 64 bits). Using this information, we determined that it corresponds to a MAC address whose vendor ID is assigned to Brocade (formerly Foundry Networks), a maker of SAN and network equipment. Assuming the MAC is not spoofed, this may indicate a faulty router. Even though it contributes a large volume of packets to background radiation, as the address is link-local and not allocated, we have no way to determine the operator to contact about possible remediation or root cause determination.

In total, we observe 205 link-local addresses in our background radiation data and over 605 in our overall three-month data, indicating that this is not a single occurrence. Likewise, we found 63 sources in our dark data and 1,678 overall sourcing packets from "unique-local" address space, which is analogous to RFC 1918 private address space and should not be globally routed [19].

These results suggests that current configurations on at least some Internet backbone and edge routers are allowing local IPv6 addresses to be globally visible, which is potentially dangerous, since it can allow a remote host to appear local to another host on the opposite end of the Internet.

### 4.6.3 Worm Activity

One of the traffic artifacts we were interested in was worm activity in IPv6. In particular, even randomly scanning worms should be visible to some degree in our data, if they exist, due to the sheer size and duration of our study. We would expect that even with a single scanning host using a slow scanning rate, such as 10 packets per second, we could expect to observe at least one packet within our four /12 blocks with 99.999% probability in 19.6 minutes [30]. Even though there had been no reports of any IPv6-capable variants of the major worms in IPv4, we decided to look for early signs of worm activity on two ports used by popular worms in IPv4.

We focused on UDP/1434 and TCP/445. These ports are used by Microsoft SQL Server and the Direct Hosted SMB protocol, respectively, and, in IPv4, they are exploited by worms such as SQL Slammer via UDP/1434, and, for example, Conficker and Sasser via TCP/445. The Slammer worm continues to be quite active in IPv4 despite the passage of a decade since the first outbreak [22]. Conficker scanning is also still highly prevalent, ranking it as the second most detected worm in the second half of 2012 [28, 37]. Validating some of this in our own work, the background radiation from the IPv4 network telescope we report on in Section 4.5 revealed over 99k UDP 1434 packets over the course of a week. We were able to positively confirm these as Slammer activity by its payload signature [36].

In our analysis of the IPv6 dark data, we noticed some small amounts of traffic on these two ports. However, upon closer inspection the data showed no signs of worm activity on either port. Our dark data contained just 18 packets to UDP port 1434, and it did not appear to be Slammer. The larger complete collected dataset also revealed nothing implicating worm activity on this port. Similarly, though we also observed some activity on port 445 (88k packets, involving a total of 92 unique IPs in the dark data), closer observations revealed that this traffic consisted of conversations between a single pair or a handful of IP address—not a typical scanning pattern. Likewise, expanding the search to the complete set of collected packets showed no indications of any scanning on these ports.

Overall, *we found no evidence of broad scanning nor prevalent malicious traffic in our data, a sharp departure from IPv4.*

### 4.6.4 NTP/BGP Services

Interestingly, we are able to find data destined to critical services such as NTP and BGP in our datasets. We find NTP traffic from over 62k unique IP addresses in the background radiation data. There are just 28 unique source IPs to 62,395 unique destinations among the packets we captured. Of the 28 sources, six are in 6to4 space (2002::/16), while the remaining are from 12 unique prefixes originated by 12 ASes—one Brazilian, one Canadian, one Mexican, and nine U.S., including two large ISPs. In B.Dark, the port used by NTP is the second most common source and destination UDP port, suggesting possible prevalent disruption of NTP activity by IPv6 misconfiguration we observe (although we note that clients often configure several NTP servers for resiliency, possibly prolonging the time to detection of the misconfiguration).

We also find a significant amount of BGP traffic in the background radiation data. In B.Dark, we see BGP packets between 338 unique IPv6 addresses. Examining the IPs of the 150 unique packet sources, we found that ten were clearly invalid, with addresses beginning with 6001::/16, a001::/16, or 1::/16, none of which fall under the global unicast address space (2000::/3); 26 of the sources were out of unrouted space, and the remaining 114 sources were originated from eight unique prefixes advertised by seven unique origin ASes: four US, two German, and one from the Philippines. These included four telecom operators and one very large global Internet search provider. In the one-day dataset, BGP traffic is the top contributor to RIPE's TCP traffic, comprising over 31% of the RIPE TCP packets, and involving eight IPs in six unique /32 blocks. The significance of BGP traffic among background radiation is that it indicates brokenness on the Internet's control plane itself, and 346 routers (if we assume no aliasing) with sources in several ASes shows that the problem is not isolated.

# 5. BEYOND BACKGROUND RADIATION

Although we used a covering prefix methodology in order to maximize our visibility, the data that we ended up collecting greatly exceeded the traditional background radiation (dark) data that we initially set out to study. We focused on deeply characterizing the dark data in section 4 because it constitutes the first large and long-term glimpse into global background radiation in IPv6, but we now turn to an analysis of this new non-traditional background radiation data, which constitutes the majority (≈95%) of the packets we captured. In the subsections that follow, we examine each of the other three types of packets: allocated/routed (AR), allocated/unrouted (AU), and unallocated/routed (UR); these comprise 35%, 60%, and <0.01% of the packets we captured, respectively.

## 5.1 Allocated/Routed (AR) Packets

As previously shown in table 4, 34.5% of the packets collected in dataset A and 36.0% of those collected in dataset B were destined to *allocated* and *routed* prefixes. Recall that, because we ended up receiving these packets, it is necessarily the case that a more-specific route to that prefix must not have been globally available at the time of the packet—at least not available between the packet's source and Merit's upstreams (AT&T and Hurricane Electric). Indeed, for several sample days we performed finer-grained analysis comparing the precise time stamp of each packet and the state of the global and local (Merit's) routing tables in order to validate that this was the case—i.e., the longest-matching-prefix routing rule was being followed. Packets only came to us when, at that instant, there was no known more-specific route, not to Merit's upstreams or not globally. In this section we first present a general examination of this allocated/routed (AR) traffic, followed by a description of a routing analysis we conducted to identify characteristics of this subset of IPv6 prefixes.

### 5.1.1 *Traffic Characterization*

Table 4 provides some high-level statistics for our AR data subset, which includes the number of packets, unique destinations, and TCP payload volume of the data in this category. Overall, we observe 4.32B packets in the three-month dataset. Of these, we categorized 1.56B packets as allocated/routed and these were directed to over 1.5M unique destination IP addresses. Below we highlight two of the key features that differentiate this data subset from the traditional background radiation. Recall that, since routes to prefixes in this category are temporarily or regionally available, *this category is more likely to be "normal" IPv6 traffic that is briefly disrupted, rather than traffic that permanently fails, as in the other categories.*

In terms of TCP flags, we find that, overall, SYNACK (76%) dominates SYN (22%) packets in the three-month dataset, B.AR. In the 24-hour collection, A.AR, it is SYN with 76%, and SYNACK at 5%. However, we note that in the A dataset, as shown earlier in Table 3, TCP packets are dominated by RIPE, which has a high percentage of SYN (78%) and very low SYNACK (just 2%). Recall that in the dark data, shown in Figure 2b, SYN packets generally dominated, i.e., connection attempts were more common. The general dominance here of SYNACK, instead suggests replies to SYN packets from these networks—replies that are unable to find a path back to the source of the antecedent packet due to the lack of a valid route.

TCP source ports are dominated by 443 and 80, at 22% and 14%, with port 5528 at 2% and all others (the next 6 being 25, 993, 995, 587, 110, and 22) below 1% of traffic. This indicates a large number of responses from web servers in the AR traffic. Notable also is the absence of port 53 in the top ten, which was the most common in the dark data. In the opposite direction, TCP destination ports are dominated by port 80 (10%) with other ports in the top 10 all under 1%. In descending order, they are: 113, 179, 22, 25, 51413, 53, 8008, 443, and 11171. This suggests some web traffic, along with client connections to services in AR nets. The presence of BGP (TCP/179 is a destination of 0.2% of packets), just as in the dark data, shows some control plane misconfiguration. In the RIPE data in dataset A.AR, the lists are similar, with 80 and 443 topping the port list in both directions. Overall, the port and protocol AR data, along with the high presence of SYNACK packets, fits the profile of client networks being prominent among AR networks.

### 5.1.2 *AR Network Prefix Properties*

To better explain this large fraction of our traffic, we sought to more closely characterize the networks in the AR set of destinations. Through our analysis (as described in 3.5) we were able to identify 311 AR networks in the dataset A and 1,669 in dataset B. Figure 5 shows the cumulative distribution function of these prefixes according to their contribution of packets to the AR data. We note that just one prefix is responsible for over 46% of the packets in A and just one (different) prefix for 21% of the packets in B, and that the top 10 prefixes account for 90% and 71% of the packets, respectively.

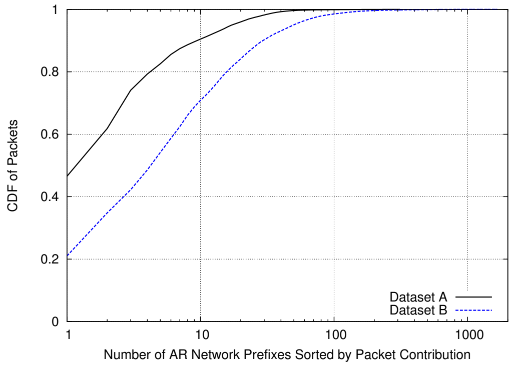


**Figure 5: Cumulative distribution of contribution of packets by 311 AR nets in dataset A and 1,669 AR nets in dataset B.**

**Table 11: Three properties of AR destination prefixes and their origin ASes as well as averages for the Internet at large (IPv6 and IPv4). All numbers are averages over the dataset, except for withdrawal events, which are averages per 24 hours.**

| | Dataset A | | | Dataset B | | |
|---|---|---|---|---|---|---|
| | **IPv4** | **IPv6** | **A.AR** | **IPv4** | **IPv6** | **B.AR** |
| **Route Analysis (over entire dataset)** | | | | | | |
| **Networks** | 485,233 | 12,841 | 311 | 502,483 | 13,462 | 1,669 |
| **R.V. Peers** | 95.22 | 65.67 | 28.68 | 102.75 | 68.34 | 65.48 |
| **Withdrawals** | 0.12 | 0.65 | 8.22 | 0.09 | 0.33 | 0.55 |
| **K-Core Decomposition Analysis (First Day of Dataset)** | | | | | | |
| **Origin ASes** | 36,559 | 6,511 | 247 | 36,651 | 6,586 | 1,018 |
| **AS Coreness** | 2.23 | 9.87 | 8.94 | 2.21 | 9.23 | 6.62 |

We hypothesized that these networks may differ systematically from the overall pool of IPv6 networks. To study this, we examined several properties of these prefixes and their origin autonomous systems (AS). For each prefix seen in the data, we tabulated the number of Route Views peers that had routes to that prefix. This is a measure of *prefix propagation*. We also counted the number of withdrawal events (which we define as the number of withdrawal messages divided by the number of peers) for each prefix seen. This is a measure of *stability*.

Lastly, for each sample day, we examined the *coreness* [3, 16] of all origin ASes. In graph theory, a $k$-core is a maximal subgraph in which every vertex has at least degree $k$. The coreness of a vertex $v$ indicates the maximum value of $N$ for which $v$ is in the $N$-core. As used in studying routing topology, coreness is a measure of AS *connectivity*. The coreness values for each origin AS were obtained via an analysis of BGP table snapshots collected from Route Views and from RIPE RIS on the first day of each dataset period.

For each of the three measures, we tabulated the overall averages for all Internet IPv4 and IPv6 prefixes, separately, as well as the values for just our list of seen AR prefixes in each sample. Table 11 summarizes our results.

*Propagation.* As can be seen, there appear to be differences between the pool of AR prefixes and the overall pool of IPv6 prefixes according to our measures. The average count of Route Views collector BGP peers advertising paths to the seen AR prefixes is about half the count of the average IPv6 prefix in the one-day sample (A), though much closer (65.48 versus 68.34) in the three-month sample (B). As with the stability measure discussed next, the difference between the three-month dataset and the one-day dataset likely has to do with the larger aggregated pool of AR prefixes over the three month period. Even a single captured packet destined to a mostly stable prefix due to a brief withdrawal would land it in our AR list. Thus, the longer time period is likely to include more stable routes than period A. We have confidence in these numbers as we also conducted these types of analyses for three single-day subsamples of the B time period and obtained very similar results to the single-day numbers for A presented here. The difference between AR prefix and overall IPv6 prefix peer counts suggests that these *AR prefixes have weaker propagation* in the global BGP routing table.

*Stability.* Next, examining the withdrawal events, we see that the overall pool of IPv6 prefixes sees an average of 0.65 withdrawal events per prefix per day for the single-day dataset, A, and 0.33 for the three-month, B. Here, we see a stark difference, as the AR prefixes exhibit 8.22 withdrawals per day in A and 0.55 withdrawal events in B. Even the latter difference is meaningful, since it indicates that the average AR prefix has 66% more withdrawals per day than the average IPv6 prefix (at 0.33 withdrawals). This suggests that AR prefixes are *much less stable* than average networks.

Through a fine-grained analysis of timestamps, mentioned above, we initially observed that some AR networks ought to have received packets that, instead, we received. We say "ought to" as routes to these networks were available in some part of global BGP, based on Route Views data. The stability and connectivity measures here explain why we received these packets in spite of their intended routes being seen by at least some Route Views peers: the routes must have been unavailable to Merit's providers. Our deeper analysis of Merit's BGP perspective confirmed that Merit's upstreams did *not* have more specific routes for the given weakly-routed prefixes at the instant each packet arrived. *Operators can mitigate lost traffic in this category via route monitoring, such as by consulting Route Views or an alert service.*

*Connectivity.* Coreness is utilized to assess the connectivity and centrality of a network. Intuitively, high coreness indicates a better-connected AS, such as a large ISP. Typical values for all IPv6-enabled ASes range from 0 (stub networks) to 70 (the best-connected ISPs). The measured coreness of 8.94 in dataset A (versus 9.87 for all IPv6 prefixes) and 6.62 for the three-month sample suggest that *the origin of these unstable and weakly-advertised prefixes is more likely to be a small network.* This is to be expected, since smaller networks might be less likely to be managed as rigorously as large Internet service providers, whose main business is the network.

### *5.1.3 IPv6 Prefix Comparison to IPv4*

Compared to the overall pool of advertised IPv6 prefixes, we found that AR prefixes had poorer propagation, were less stable, and were originated by less well-connected autonomous systems. As IPv6 is still at an early deployment phase, we expect IPv6 prefixes, as a whole, to have poor routing. To perform a systematic juxtaposition of IPv4 against IPv6, we repeated the analyses described above to obtain the corresponding IPv4 metrics.

Table 11 also depicts these measures applied to the entire pool of IPv4 prefixes. We first observe that the IPv4 networks are, in general, better-connected to Route Views monitors, with average peer counts of 95.22 and 102.75 for the A and B samples, respectively. Examining the average withdrawals, we see that IPv4 prefixes are much more stable, exhibiting just 0.12 (A) and 0.09 (B) withdrawal events per prefix per day. Finally, we see that the average coreness of IPv4 prefixes is much lower than that of IPv6. This means that the average IPv4 prefix is originated by a smaller network than the average IPv6 prefix. This is not surprising, since it has been shown that IPv6 deployment is more prevalent at the core of the Internet than at the edges [13], and, both by definition and by necessity, the core is highly connected.

These findings suggest that, indeed, IPv6 routing is not as mature as IPv4, and that the AR networks are an even less mature subset of the IPv6 pool. This result is intuitive, given the early stage of IPv6 deployment, the fact that it only carries a small fraction (≈0.2%) of Internet traffic [21], that operators generally have much less experience with the new protocol, and that both hardware and software support for the new protocol is still not on par with IPv4 support.

## 5.2 Allocated/Unrouted (AU) Packets

As table 4 shows, 60.4% of dataset A and 59.2% of B packets are destined for addresses that match *allocated but unrouted* prefixes. This is interesting, as it suggests traffic to networks that were assigned and configured but which were not intended to be globally reachable—perhaps because they were intended for internal (i.e. "private") organizational use, in spite of existence of the *unique local* prefix, fc00::/7, analogous to IPv4 RFC 1918 space, for this purpose [19, 32]. Indeed, traffic patterns of the two largest contributors we found in this category, were confirmed examples of this.

Together with the Allocated/Routed (AR) packets, this category of traffic constitutes nearly 95% of packets we saw. This high-level finding, that a majority of collected traffic via a covering prefix network telescope belongs to allocated networks, is consistent with an earlier experiment with just APNIC's /12 reported by Geoff Huston [20]. Here, we highlight some of the largest contributors we discovered in this subset of the data.

We observed periodic spikes in the overall volume of traffic in the ARIN dataset. These large jumps, resulting in a one or two order of magnitude spike in ARIN traffic, occurred daily at around noon UTC. Upon investigation, we found that they were largely composed of DNS replies. All of these packets had their destinations set to the same value, which we omit here to preserve organizational privacy. In total, we received over 444M packets (over 423M of which were DNS responses) destined to this one IP. It is the top contributor and accounts for over 17% of all packets in the AU dataset. The address is in a network block that has been allocated to a managed hosting company but is not seen in any routing tables visible to any Route Views monitors.

We contacted operators at this organization to inform them of the traffic and learn the cause. It was quickly determined that the source of the traffic was a misconfigured DNS server. The server was assigned a (globally unique) address meant by the organization for internal use, but it was using this address as the source of external packets it sent. Since this address was from a prefix that was not globally advertised in BGP, the replies to that DNS server all ended up being routed, via our covering prefix, to our collector. This misconfigured server was not previously noticed by its operators because it was part of a cluster of several resolvers, which *were* correctly configured. After the company began advertising the network block in which this server was addressed in mid-August 2013, we stopped receiving these packets, and our total ARIN traffic, in bytes, fell by about half.

Another large contributor to the AU dataset is traffic destined to a network block allocated to the R&D unit of a large U.S. wireless phone provider. We received 1.1Bn packets destined to 216 destinations in one prefix allocated to this organization. This constitutes nearly 44% of the packets in the AU dataset. There were over 6M unique IPs sending traffic to these 216 destinations, and nearly 6M were from the same /32 block as the destinations. Most of the sources were from sub-blocks of the /32 that were publicly routed. This strongly suggested that the prefix in question was being used internally but that some routing misconfiguration was causing traffic from other prefixes within that organization to be misdirected out to the Internet instead of toward that internal network. We established contact with the company and they confirmed that this address space is used internally and not meant to be globally reachable. At publication deadline, they were still investigating why these internal hosts were sending their traffic to our collector instead of to the local private network. In general, operators can avoid leaking such internal traffic by either using unique local addressing internally or by placing access lists on edge routers to either block or log such unintended address use.

## 5.3 Unallocated/Routed (UR) Packets

The third and final category, accounting for less than 0.01% of packets in both the A and B time periods, is those whose destination address matching prefixes that were not allocated by an RIR but which did have routes in the BGP routing table at least temporarily during our data collection period. Again, recall that, these prefixes were not globally reachable during the instant we captured a given packet, otherwise it would not be routed to us. We suspected that these unallocated but either partially- or temporarily-routed prefixes may be either the result of RIRs withdrawing allocations while operators continued to use the address space, the result of experimentally-advertised prefixes, or the result of misconfiguration.

In dataset period A, there were two prefixes in this category, 2a02:510::/32 and 2606:8900::/32, and there were also two during period B: 2607:fd20::/32 and 2406:7a00::/31. Each of these four prefixes was unallocated at the end of our data collection windows. However, the first of the four was in allocation for several years prior to our collection, the second was allocated a few days after the collection period, the third was allocated for part of our B collection period then returned, and the last was never allocated (but a prefix one bit longer was).

In each case, these prefixes were known to Route Views by only a single peer router (versus ≈75 for the overall average IPv6 prefix, as discussed in section 3.2), suggesting very limited announcements, likely for testing or due to misconfiguration. Because of this, the traffic in this category most closely resembles the unallocated and unrouted (i.e. background radiation) traffic.

Focusing on the packets, we find that ICMPv6 comprised 94.5 and 98.1% of the traffic in this category in datasets A.UR and B.UR, respectively. Upon further examination, we discovered that the vast majority of packets were due to traffic from hosts associated with researchers. In dataset B, for example, 83.5% of packets came from what we confirmed with CAIDA to be their experimental hosts. Another 8.8% of the packets were confirmed by BBN Technologies as hosts that are part of their experiments. A third source, with 2.8% of these packets was confirmed as belonging to another research organization we contacted, in Canada. *In total, at least 95.1% of the packets in the B.UR (and 97.5% in the A.UR) dataset were the result of confirmed researcher activity.*

## 5.4 Near Misses

When we modified our initial RIPE NCC /12 announcement into a /13 and a /14, eliminating the 25% of the address space from which RIPE was making customer allocations, the volume of RIPE traffic we saw dropped disproportionately—by three orders of magnitude. We considered whether this was representative of a general case, i.e., if the majority of network telescope traffic we received was clustered near allocated address space. We discussed most of the answer to this above by showing that, after our allocation and routing analysis, around 95% of the captured traffic was, indeed, destined for allocated prefixes (the AU and AR data subsets). However, what about traffic to unallocated-and-unrouted space?

We hypothesized that if misconfiguration was a large source of the traffic to unallocated networks, then the target destinations might be addresses that are lexically close to existing, active prefixes (i.e., "fat finger" errors). To examine this systematically, we conducted a Levenshtein distance [27] analysis of each packet's destination in fully-expanded ASCII format, comparing it to all known routed networks (up to the prefix length). An analysis of three day-long samples of destination IPs of packets we captured (all of dataset A, as well as all packets on the first day of each month of dataset B) yielded the somewhat unexpected result that *every destination address had some minimal edit distance of at most 2*. Between 39% and 81% of packets, depending on the sample, were within one hex character change away from an existing routed prefix, and the rest were within two. Since our other analyses showed that such a large percentage of traffic is for allocated networks, this should not be too surprising, as routed prefixes are often part of or near other allocated addresses. However, that this held even for the pure "dark" traffic to unrouted-and-unallocated space was unexpected. While we can't confirm this, it suggests that one explanation for the dark traffic could simply be typos, as a single character change would bring each packet we saw under a legitimate routed prefix.

# 6. CONCLUSION

In this paper, we presented results from our broad observation of Internet pollution on the IPv6-enabled Internet. To the best of our knowledge, this study, which relies on five large covering prefixes, is the first of its kind in terms of visibility.

Our results show that misconfiguration is prevalent in the IPv6 network, although concurrent measurements we made in IPv4 indicate that the shear volume of background radiation is minuscule (500-fold less) in IPv6 despite a vastly larger covered address space. In fact, in the IPv6 traditional background radiation data, a single block with just 66 unique addresses was responsible for 92% of the packets—nearly all were ICMPv6 echo requests or replies. Happily, we found no evidence of prevalent malicious traffic in the captured IPv6 data due to worms, scanning, or backscatter, which are all common in IPv4 background radiation. We continue our collection, in an effort to understand the long-term trends as well as to identify particular configuration and stability findings that can be shared with the broader Internet operations community.

We found that only about five percent of the packets we captured were destined for unallocated and unrouted address space, the type of addresses that network telescopes traditionally monitor. Thus, our use of a covering prefix is a departure from most network telescope studies in the past. It provided unique visibility into routing and configuration errors previously hidden from this type of study—i.e., those affecting allocated prefixes. Nearly 60% of packets we captured were destined to allocated but unrouted destinations, likely due to traffic sourced by networks meant for internal use being "leaked" into the Internet; this was the case for the two largest contributors to this category, which we contacted, accounting for 61% of packets. Another 36% of packets were destined to allocated space that had, at least briefly or sparsely, advertised routes. We examined updates for these prefixes and found them to be less stable and advertised than the average IPv6 prefix, which is itself less stable and less well-advertised than the average IPv4 prefix. We received these packets due to unavailable routes. The final category of data was mostly explained by researcher traffic, which we confirmed by contacting several organizations. In aggregate, we identified the cause of a majority of all captured packets. We aim to continue covering-prefix-based network telescope studies, including in IPv4 space, due to the increased insight they provide.

## Acknowledgments

This work was supported in part by the Department of Homeland Security Science and Technology Directorate under contract numbers D08PC75388, FA8750-12-2-0314, and FA8750-12-2-0235; the National Science Foundation (NSF) under contract numbers CNS

1111699, CNS 091639, CNS 08311174, CNS 0751116, CNS 1330142, and CNS 1255153; and the Department of the Navy under contract N000.14-09-1-1042. We are also grateful for help from Rick Reinecke at Merit, from our shepherd, and from the anonymous IMC reviewers who provided valuable manuscript feedback.

## 7. REFERENCES

[1] E. Aben, A. King, K. Benson, Y. Hyun, A. Dainotti, and K. Claffy. Lessons learned by "measuring" the Internet during/after the Sandy storm. In *In Proceedings of FCC Workshop on Network Resiliency 2013*, Feb. 2013.

[2] Alexa Internet, Incorporated. Top 1,000,000 Sites. http://s3.amazonaws.com/alexa-static/top-1m.csv.zip.

[3] J. I. Alvarez-Hamelin, L. Dall'Asta, A. Barrat, and A. Vespignani. k-core decomposition: a tool for the visualization of large scale networks. *CoRR*, abs/cs/0504107, 2005.

[4] M. Bailey, E. Cooke, F. Jahanian, A. Myrick, and S. Sinha. Practical Darknet Measurement. In *Proceedings of the 40th Annual Conference on Information Sciences and Systems (CISS '06)*, pages 1496–1501, Princeton, New Jersey, USA, Mar. 2006.

[5] M. Bailey, E. Cooke, F. Jahanian, N. Provos, K. Rosaen, and D. Watson. Data Reduction for the Scalable Automated Analysis of Distributed Darknet Traffic. *Proceedings of the USENIX/ACM Internet Measurement Conference*, Oct. 2005.

[6] P. Barford, R. Nowak, R. Willett, and V. Yegneswaran. Toward a Model for Source Address of Internet Background Radiation. In *Passive and Active Measurement Conference (PAM)*, 2006.

[7] N. Brownlee. One-Way Traffic Monitoring with iatmon. In N. Taft and F. Ricciato, editors, *PAM*, volume 7192 of *Lecture Notes in Computer Science*, pages 179–188. Springer, 2012.

[8] E. Cooke, M. Bailey, Z. M. Mao, D. Watson, and F. Jahanian. Toward understanding distributed blackhole placement. In *Proceedings of the 2004 ACM Workshop on Rapid Malcode (WORM-04)*, New York, Oct 2004. ACM Press.

[9] E. Cooke, M. Bailey, D. Watson, F. Jahanian, and J. Nazario. The Internet motion sensor: A distributed global scoped Internet threat monitoring system. Technical Report CSE-TR-491-04, University of Michigan, Electrical Engineering and Computer Science, July 2004.

[10] A. Dainotti, A. King, K. Claffy, F. Papale, and A. Pescapè. Analysis of a "/0" Stealth Scan from a Botnet. In *Internet Measurement Conference (IMC)*, Nov 2012.

[11] 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. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, IMC '11, pages 1–18, New York, NY, USA, 2011. ACM.

[12] C. Deccio. Turning Down the Lights: Darknet Deployment Lessons Learned. Technical report, Sandia National Laboratories, 2012.

[13] A. Dhamdhere, M. Luckie, B. Huffaker, K. Claffy, A. Elmokashfi, and E. Aben. Measuring the Deployment of IPv6: Topology, Routing and Performance. In *Proceedings of the 12th ACM SIGCOMM Conference on Internet Measurement (IMC 2012)*.

[14] M. Ford, J. Stevens, and J. Ronan. Initial Results from an IPv6 Darknet. In *International Conference on Internet Surveillance and Protection*, 2006.

[15] E. Glatz and X. Dimitropoulos. Classifying internet one-way traffic. pages 417–418, 2012.

[16] G. Gürsun, N. Ruchansky, E. Terzi, and M. Crovella. Inferring visibility: who's (not) talking to whom? *SIGCOMM Comput. Commun. Rev.*, 2012.

[17] W. Harrop and G. Armitage. Defining and evaluating greynets (sparse darknets). In *Proceedings of the The IEEE Conference on Local Computer Networks 30th Anniversary*, LCN '05, pages 344–350, Washington, DC, USA, 2005. IEEE Computer Society.

[18] R. Hinden and S. Deering. RFC 4291: IP Version 6 Addressing Architecture, 2006.

[19] R. Hinden and B. Haberman. RFC 4193: Unique Local IPv6 Unicast Addresses, 2005.

[20] G. Huston. IPv6 Background Radiation. Technical report, 2012. Slides of a talk given at DUST 2012 – The 1st International Workshop on Darkspace and UnSolicited Traffic Analysis, May 14–15, San Diego, California.

[21] J. Czyz, M. Allman, J. Zhang, S. Iekel-Johnson, E. Osterweil, and M. Bailey. Measuring IPv6 Adoption. ICSI Technical Report TR-13-004, August 2013.

[22] J. Jarmoc. SQL Slammer âĂŞ 10 years later. http://www.secureworks.com/resources/blog/research/general-sql-slammer-10-years-later/, Mar. 2013.

[23] M. Karir, G. Huston, G. Michaelson, and M. Bailey. Understanding ipv6 populations in the wild. In *Passive and Active Measurement (PAM '13)*. 2013.

[24] E. Karpilovsky, A. Gerber, D. Pei, J. Rexford, and A. Shaikh. Quantifying the Extent of IPv6 Deployment. In *Proceedings of the 10th International Conference on Passive and Active Network Measurement*, 2009.

[25] A. King. Syria disappears from the Internet. http://blog.caida.org/best_available_data/2012/12/05/syria-disappears-from-the-internet/, 2012.

[26] C. Kreibich, N. Weaver, B. Nechaev, and V. Paxson. Netalyzr: illuminating the edge network. In *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*, IMC '10, pages 246–259, New York, NY, USA, 2010. ACM.

[27] V. I. Levenshtein. Binary codes capable of correcting deletions, insertions and reversals. *Sov. Phys. Dokl.*, 6:707–710, 1966.

[28] Microsoft. Microsoft Security Intelligence Report - July–December 2012. http://download.microsoft.com/download/E/0/F/E0F59BE7-E553-4888-9220-1C79CBD14B4F/Microsoft_Security_Intelligence_Report_Volume_14_English.pdf, 2013.

[29] D. Mills, J. Martin, J. Burbank, and W. Kasch. Network Time Protocol Version 4: Protocol and Algorithms Specification. RFC 5905 (Proposed Standard), June 2010.

[30] D. Moore, C. Shannon, G. M. Voelker, and S. Savage. Network telescopes. Technical Report CS2004-0795, UC San Diego, July 2004.

[31] R. Pang, V. Yegneswaran, P. Barford, V. Paxson, and L. Peterson. Characteristics of Internet background radiation. In *Proceedings of the 4th ACM SIGCOMM conference on Internet measurement*, pages 27–40. ACM Press, 2004.

[32] Y. Rekhter, B. Moskowitz, D. Karrenberg, G. J. de Groot, and E. Lear. Address Allocation for Private Internets, 1996.

[33] RIPE NCC. Routing Information Service (RIS). http://www.ripe.net/ris/.

[34] N. Sarrar, G. Maier, B. Ager, R. Sommer, and S. Uhlig. Investigating IPv6 Traffic - What Happened at the World IPv6 Day? In *Proceedings of the 13th International Conference on Passive and Active Network Measurement*, 2012.

[35] University of Oregon. Route Views project. http://www.routeviews.org/.

[36] VigilantMinds. MS-SQL Slammer Signature. http://seclists.org/snort/2003/q1/871, Jan. 2003.

[37] S. Walls and M. Karir. Internet pollution - part 2. Presented at the 51st North American Network Operators Groop (NANOG51) Meeting in Miami, Florida, Feb. 2011.

[38] E. Wustrow, M. Karir, M. Bailey, F. Jahanian, and G. Houston. Internet Background Radiation Revisited. In *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement (IMC '10)*, Melbourne, Australia, November 2010.

[39] V. Yegneswaran, P. Barford, and D. Plonka. On the design and use of Internet sinks for network abuse monitoring. In *Recent Advances in Intrusion Detection—Proceedings of the 7th International Symposium (RAID)*, Sophia Antipolis, French Riviera, France, 2004.

[40] G. Zhang, B. Quoitin, and S. Zhou. Phase changes in the evolution of the IPv4 and IPv6 AS-Level Internet topologies. *Computer Communications*, 34(5), Apr. 2011.