Estimating Internet Address Space Usage through Passive Measurements

Alberto Dainotti, Karyn Benson, Alistair King, kc claffy
CAIDA, UC San Diego
La Jolla, California, USA
{alberto,karyn,alistair,kc}@caida.org

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

Eduard Glatz, Xenofontas Dimitropoulos
ETH Zurich
Zurich, Switzerland
{eglatz,fontas}@tik.ee.ethz.ch

ABSTRACT

One challenge in understanding the evolution of Internet infrastructure is the lack of systematic mechanisms for monitoring the extent to which allocated IP addresses are actually used. Address utilization has been monitored via actively scanning the entire IPv4 address space. We evaluate the potential to leverage passive network traffic measurements in addition to or instead of active probing. Passive traffic measurements introduce no network traffic overhead, do not rely on unfiltered responses to probing, and could potentially apply to IPv6 as well. We investigate two challenges in using passive traffic for address utilization inference: the limited visibility of a single observation point; and the presence of spoofed IP addresses in packets that can distort results by implying faked addresses are active. We propose a methodology for removing such spoofed traffic on both darknets and live networks, which yields results comparable to inferences made from active probing. Our preliminary analysis reveals a number of promising findings, including novel insight into the usage of the IPv4 address space that would expand with additional vantage points.

Categories and Subject Descriptors

C.2.3 [**Network Operations**]: Network monitoring;
C.2.5 [**Local and Wide-Area Networks**]: Internet

Keywords

Passive measurements; IPv4 address space; Internet address space; Darknet; Network telescope; Spoofed traffic; Internet census

1. INTRODUCTION AND MOTIVATION

On February 3, 2011, the Internet Assigned Numbers Authority (IANA) allocated its last set of available IPv4 addresses to Regional Internet Registries (RIR), a historic turning point for the Internet. The address pools of the European and Asian-Pacific RIRs were exhausted in 2012; the other RIRs will likely run out within the next few years. Although IPv4 address scarcity is now a reality, so is the fact that allocated addresses are often heavily under-utilized. One challenge in managing Internet address space (both IPv4 and IPv6) is the lack of reliable mechanisms to monitor actual utilization of addresses. Macroscopic measurement of patterns in IPv4 address utilization also reveals insights into Internet growth, including to what extent NAT and IPv6 deployment are reducing the pressure on (and demand for) IPv4 address space.

To our knowledge, the only previous scientific work mapping actual utilization of IPv4 addresses is ISI's Internet Census project [10] (ISI Census, in the following), which periodically sends ICMP echo requests to every single IPv4 address (excluding private and multicast addresses) to track the active IP address population. This approach has four primary limitations: significant probing overhead; potential to offend probing targets (i.e., the entire Internet) who may request not to be probed or even blacklist probing addresses; inaccuracies due to the fact that many networks either filter out ICMP echo requests or respond to them on behalf of other IP addresses, and inability to scale for use in a future IPv6 census.

We investigate the potential for passive network traffic measurements to inform or even substitute for active methods of measuring address space utilization. Passive measurements introduce no network traffic overhead, do not rely on unfiltered responses to probing, and could potentially apply to IPv6 as well. On the other hand, a passive-measurement approach to address utilization inference has two daunting challenges which we explore in this paper: the limited visibility of any single vantage point of traffic; and the presence of spoofed IP addresses in packets that can significantly distort results by implying faked addresses are active. We analyze two types of passive traffic data: (i) Internet Background Radiation (IBR) packet traffic[1] [25] captured by darknets (also known as network telescopes); (ii) traffic (net)flow summaries in operational networks. We develop and evaluate techniques to identify and remove likely spoofed packets from both darknet (unidirectional) and two-way traffic data. Our **contributions** include:

[1]IBR is generated primarily by malware.

- We demonstrate that, surprisingly, passive traffic measurements, including both packet-level captures from darknets and two-way network flow summary records from operational networks, can reveal substantial insight into Internet address utilization.
- We introduce and validate heuristics for detecting IP address spoofing in captured network traffic data, thus mitigating their impact on our inferences.
- We use our traffic data sets to show that combining passive with active measurements can reduce probing (by 38.5%) and find an additional $\approx 450K$ active /24s not detected as active by ICMP-based Internet census measurement.
- Our initial results reveal new macroscopic insights into Internet address space utilization, such as increasing activity within large legacy address blocks; additional vantage points would likely illuminate activity in other regions of Internet address space.
- We publish comprehensive maps of IPv4 address space utilization, at a /24 granularity, for 2012 [7].

2. FRAMING THE PROBLEM

To investigate the potential of passive measurements to provide a census-like snapshot of IP address space usage, we consider four factors: finding appropriate traffic measurement locations; selecting an observation granularity; identifying and removing traffic with spoofed IP source addresses; and quantitatively evaluating our algorithm including validating against what ground truth data we can gather.

Observation granularity. We analyze the usage of the IPv4 address space with /24 address-block granularity (/24 blocks, in the following), i.e., we consider a /24 block as either *active* (or *inactive*) if we observe traffic from at least one (or exactly zero) IP address from that address block. There is no universal IP address segment boundary space (due to sub-netting and varying size of administrative domains), but using a /24 granularity mitigates the effects of dynamic but temporary IP address assignment (e.g., DHCP), as well as having an intuitive relationship with both routing operations and address allocation policy.

Measurement location. We devise techniques for two types of passive measurement data: bidirectional netflow data from all nodes in a large live research network (SWITCH, in Switzerland), and raw unidirectional packet traffic data from two large darknets. Section 3 describes these data sets, and others we used to validate our inferences.

Identifying and removing spoofed traffic. Packets with source IPs other than those assigned to the sending host, i.e., *spoofed* traffic, will skew our address utilization results. Source IP addresses may be spoofed for a variety of reasons, typically to prevent attribution of an attack and avoid being caught [6][2], but also due to transmission or programming (human) errors that induce address bit errors. Darknets are less likely to be the target of spoofed DoS attacks since there are no hosts to attack, but both dark and live networks receive unintentionally spoofed packets. Accurately identifying passively observed spoofed traffic may be impossible, but we propose heuristics to exclude large fractions of spoofed traffic in order to limit its impact on our estimates of IPv4 space usage. Section 4 describes our methodology for the two different types of traffic data.

Evaluating our approach. Finally, we assess how accurately we have distinguished between active and inactive areas of the IPv4 address space, given limited ground truth (described in Section 3). Our techniques yield four types of inferences: true positives (tp), false positives (fp), true negatives (tn) and false negatives (fn). We evaluate their performance using four standard metrics:

- *Precision* $= \frac{tp}{tp+fp}$: fraction of positives that are true positives.
- *Recall* $= \frac{tp}{tp+fn}$: fraction of "active"-labeled networks correctly reported as active.
- *True negative rate* $= \frac{tn}{tn+fp}$: fraction of "inactive"-labeled networks correctly reported as inactive.[3]
- *Accuracy* $= \frac{tp+tn}{tp+tn+fp+fn}$: fraction of correctly classified positives and negatives.

3. DATASETS

Our passive measurement data include: (i) full packet traces collected from a /8 network telescope operated by the University of California San Diego (UCSD) [24]; (ii) full packet traces collected from a ≈/8 network telescope (space covering 14M different addresses) operated by Merit Network (MERIT) [17]; and (iii) unsampled NetFlow traces collected at SWITCH, a regional academic backbone network that serves 46 single-homed universities and research institutes in Switzerland [22]; the monitored address range of SWITCH contains 2.2 million IP addresses, which correspond to a continuous block slightly larger than a /11. All datasets are from the same collection period, between July 31 and September 2, 2012.

For comparison, we use the ISI Internet Census dataset *it49c-20120731*, obtained by probing the entire IPv4 address space [12] during this same time window. Based on the suggestions by the data collector [13], we filtered this dataset to only those probes that received an ICMP echo reply, and where the source address in the reply matched the target address of the original probe. We do not consider negative replies (e.g. ICMP time exceeded) since we assume they are not reaching the target. In addition, we filtered out all probes sent to IP addresses with a least significant byte of *0* or *255*, which are typically reserved for network and broadcast addresses respectively [10] and third parties along the path may intercept and reply to such probes. Note that the ISI Census experiment was designed to report at a /32 (host) rather than /24 (subnet) granularity, but we apply the resulting data set to a /24 granularity analysis.

To establish a set of unrouted IPv4 address blocks, we take a list of BGP prefixes announced and captured by the *route-*

[2]nmap [15], a popular scanning tool, includes an option to add spoofed traffic to a scan to ambiguate the scanner's actual IP address.

[3]This metric is also known as specificity.

views2.routeviews.org [3] collector between July 31 and September 2, 2012, and assume all other address blocks are unrouted. We construct a validation "ground truth" data set from this same BGP data and the ISI Census snapshot described above. We label as "inactive" all unrouted /24 blocks (≈6.5M), assuming they should not appear in (unspoofed) traffic, and we label as "active" all the /24 blocks found responsive in the ISI census dataset (≈4.3M). This validation data set is limited by the lack of ground truth for address blocks that do not respond to probing.

4. TECHNIQUES FOR A PASSIVE CENSUS

We develop heuristic techniques to filter IP address spoofing from passive traffic measurements on both live networks (§ 4.1) and darknets (§ 4.2), and evaluate their effectiveness using the four metrics in Section 2.

4.1 Measurements from a live network

We developed a heuristic to identify used IP address blocks from traffic flow data, e.g., NetFlow records, and tested it on unsampled NetFlow records collected from all border routers of SWITCH. Our heuristic relies on the typical nature of TCP: if we see two IP addresses engaged in a two-way TCP connection, we assume neither address is spoofed. Our methodology has two steps: we first find two-way TCP connections and then remove connections with too few packets or bytes. A TCP connection is identified by the standard 5-tuple: source and destination IP addresses and port numbers and the layer-4 protocol. We did not include TCP flags or other header fields since they do not appear in our NetFlow data. To create two-directional flows, we must first reconstruct flows that NetFlow has fragmented into multiple records due to flow expiration mechanisms (used to limit the size of the flow table). Specifically, we merge flow records with identical 5-tuples within the same 10-minute interval into a single flow, as well as flows within the same or adjacent time intervals with the same 5-tuple, but with reverse values in the source and destination fields. The result is a two-way flow.

In some cases a two-way flow may still involve a spoofed IP address. For example, repling to a connection attempt containing a spoofed source IP address yields a two-way flow. For this reason, we require a TCP flow to carry a minimum number of packets in order to classify its addresses as not spoofed. We require by default at least 5 packets (a 3-packet handshake; 1 packet of payload; and typically 4 but at least 1 packets to reset or terminate a connection) and evaluate the impact of alternative values. To gain more confidence in situations that might include retransmissions (thus surpassing the packet threshold even if a TCP connection is not established), we add a minimum average packet size requirement, which implies that TCP payload is present. IP and TCP headers are 20-bytes, each may hold 40 additional bytes for options, so we evaluate multiple minimum average packet sizes between 40 and 120 bytes (on layer 3) before selecting a value for this parameter.

Next we analyze how the two parameters of our heuristic, i.e., minimum number of packets and average packet size, affect the results. Table 1 illustrates how the unique count of inferred active /24 blocks and the corresponding percentage that are actually unrouted change with different values of these two parameters. The percentage of unrouted blocks inferred as active is low, i.e., below 0.1%, for most parameter settings, while the coverage ranges between 3.3M and 4.2M /24 blocks. Relaxing the parameters increases the coverage, but infers a higher fraction of unrouted subnets as active. Based on the data in Table 1, we conservatively set the average packet size threshold of our heuristic to 80 bytes (as a smaller choice results in a steep increase in the percentage of unrouted blocks). In addition, we keep the default 5 packet requirement, since other values (except for the 2 packet value, which increases significantly the percentage of unrouted blocks), only moderately affect the two metrics[4]. Applying our heuristic reduces the number of /24 blocks we infer as active from 12.9M (without filtering IP address spoofing) to 3.6M.

4.2 Measurements from darknets

The absence of legitimate users and the reduced amount of traffic reaching a darknet (telescope) make it a convenient vantage point to collect and inspect traffic, both for privacy and logistical concerns. But darknets normally only see incoming traffic and do not respond, making bidirectional flow-based data analysis techniques inapplicable. Also, traffic reaching darknets comes from a variety of unpredictable sources (such as malware and misconfigurations at different layers of the protocol stack), so defining "normal" traffic is inherently difficult. To mitigate the effects of spoofing on darknet measurements, we focus on identifying and filtering out large portions of spoofed traffic, rather than first identifying unspoofed traffic as we do with bidirectional traffic. We build signatures for our filters by identifying suspicious traffic components, manually isolating and analyzing them, and then defining a filter to remove them. We focus on spoofed traffic that appears to originate from many sources (such as randomly spoofed traffic), which we call *large-scale* spoofing. We assume that the remainder of spoofing is not only difficult to detect without responding to received packets, but has a much smaller impact on our inferences, which we confirm at the end of this section.

In search of large-scale spoofing, we look for two distinct behaviors:

1. bursty behavior – (i) sudden spikes in the number of unique source IP addresses, unique source /24 blocks, and newly observed source IP addresses (source /24 blocks) per hour;

   (ii) the same type of events with only source addresses in unrouted network blocks (which normally should not generate traffic);

2. long-term consistent behavior: (i) we aggregate packets over the entire measurment window into traffic classes by protocol and port (when applicable) and investigate classes with many originating unrouted /24 blocks; (ii) we aggregate packets based on the least significant byte of the source address to look for inconsistencies in address utilization.

[4]Since handshake packets typically do not carry payload, the 80-byte average packet size requirement might be too strict. However, payload in SYN packets is not excluded by the TCP protocol definition [16].

	2 packets	3 packets	4 packets	5 packets	6 packets	7 packets
40 bytes	4,234,657 (2.22%)	3,948,088 (0.088%)	3,877,972 (0.062%)	3,847,184 (0.024%)	3,841,394(0.011%)	3,780,039 (0.0071%)
80 bytes	3,634,485 (0.0012%)	3,634,150 (0.00091%)	3,634,047 (0.00088%)	**3,633,905 (0.00083%)**	3,632,279 (0.00080%)	3,630,416 (0.00072%)
120 bytes	3,514,057 (0.00097%)	3,513,738 (0.00080%)	3,513,529 (0.00077%)	3,513,284 (0.00074%)	3,510,953 (0.00071%)	3,510,284 (0.00063%)

Table 1: Unique count of /24 blocks in the SWITCH dataset seen in two-way TCP flows with the given minimal packet count and minimal average packet size requirements. The figure in parentheses shows the percentage of the /24 blocks that are unrouted. In bold we highlight our selected parameter setting.

	Filter	Characterization	Num. /24s		Num. Unrouted /24s		Num. Dark /24s	
			UCSD	MERIT	UCSD	MERIT	UCSD	MERIT
General	TTL> 200 and not ICMP	Large-scale/Bursty	10,095,728	9,745,800	30,883	68,874	119,163	68,338
General	Least signif. byte src addr 0	Large-scale/Bursty	662,689	427,214	21,795	1,717	3,874	1,006
General	Least signif. byte src addr 255	Large-scale/Consistent	328,380	274,078	296	1,257	34	54
General	Protocol 0	Large-scale/Bursty	60,913	16,919	15,849	354	512	106
General	Protocol 150	Large-scale/Consistent	339	33	79	0	0	0
General	Same Src. and Dst. Addr.	Small-scale	625	1	0	0	625	1
Specific	All Specific Filters	Large-scale	1,937,886	976,153	532,711	286,684	15,916	7,506

Table 2: Summary of filtering heuristics used in darknet measurements and their impact in terms of source /24 blocks. We defined filters that captured general characteristics of spoofing, but in some cases we eliminated spoofing traffic specific to our darknets. For each general filter and the aggregate of all the specific filters, we report the total number of /24 blocks used as sources in packets captured by the darknets, as well as the number that are unrouted and dark.

For bursty traffic, we applied a simple spike-detection algorithm, with a threshold of 25% more than the average value observed over the last ten hourly time bins. We tried different values of these parameters without observing significant changes in what was detected as spoofed, since most events of interest cause large traffic variations.

Figure 1 shows that some bursty spoofing events are not visible when considering packets from all sources, but they become easily detectable when looking only at source addresses of unrouted networks. In some cases, this phenomenon is due to the nonuniform distribution of unrouted networks over the address space, e.g., the temporary popularity of some address blocks as source addresses despite little change in total number of spoofed sources.

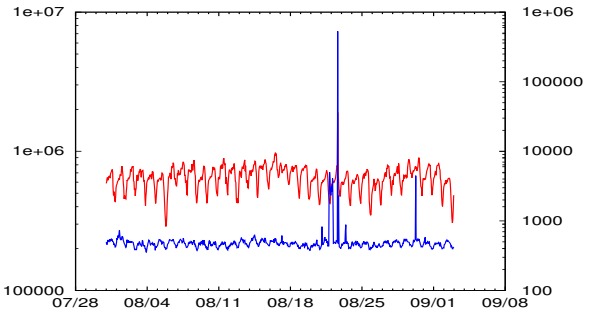

Figure 1: Routed and unrouted networks by hour (UCSD). For both darknets, we observe significant increases in the number of unrouted source networks for some hours, which we inspect to discover and exclude spoofed traffic.

To identify consistent spoofing behavior, we aggregated the entire data set by protocol and port, and then examined classes of traffic with either more than 10 unrouted /24s or a percentage of unrouted /24s greater than 0.4%. Classes with traffic below these thresholds were difficult to infer as spoofed based on traffic patterns, but these thresholds are sufficient to remove large-scale events. We also aggregated based on the least significant byte of the source address, since packets with byte 0 and byte 99 were symptomatic of several spoofing spikes.

Each spoofing behavior we identified exhibited distinctive properties, which we synthesized into a set of filtering heuristics. Table 2 lists how many /24 blocks (respectively total, unrouted-only, dark-only) originated traffic matching each heuristic. The first heuristic, based on the value of the TTL IP header field, filters out by far the largest number of /24 blocks. We found 20 spikes (11 in UCSD data and 9 in MERIT data) where a significant number of UDP packets with unrouted sources had the same destination port and TTL above 200. Our filter excludes traffic based on the large TTL since it indicates a general abnormality: most operating systems use a default a TTL of 128 or less [20] (although, several switch to a TTL of 255 when sending ICMP packets). Our heuristics based on traffic spikes filter out not only the likely spoofed traffic during the spike, but also traffic matching this filter outside of the spike.

Other significant portions of spoofed traffic use uncommon or unassigned protocols, but such behavior could also be legitimately experimental so we do not exclude traffic solely for this reason. But when many packets with an uncommon protocol appear to originate from unrouted addresses, it is more likely they are the result of bit-flips during transmission or programming errors when writing packets. We exclude packets with source address ending in .0 or .255 since traffic should not originate from these addresses (when part of a /24 subnet). We also identified small spoofing events where the source and destination were in the same darknet. Finally, the last line of Table 2 aggregates results for a set of filtering criteria specifically crafted for abnormal events observed in one of the darknets. They do not seem generally applicable, so we only report their effect of their use on our datasets.[5]

[5]An example is UDP packets from multiple sources all des-

			Number of /24 blocks (sources)	
Monitored destination	UCSD		MERIT	SWITCH-DARK
		before filtering	54210 (98.4%)	4522 (98.9%)
		after filtering	21 (0.038%)	0 (0%)
	MERIT		UCSD	SWITCH-DARK
		before filtering	57769 (91.5%)	4379 (95.7%)
		after filtering	8 (0.013%)	1 (0.022%)

Table 3: Our filtering in the darknet datasets dramatically reduces the percentage of /24 blocks erroneously inferred as active while known to be spoofed (because they appear to originate from the darknets or unused blocks of SWITCH). These blocks originally appear as up to 98.9% active; filtering lowers their inferred usage to 0.038% or less.

	Precision	Recall	True Negative Rate	Accuracy
UCSD	0.998	0.672	0.999	0.869
MERIT	0.999	0.645	0.999	0.859
SWITCH	0.999	0.756	0.999	0.903
Total	0.998	0.811	0.999	0.924

Table 4: Validation of passive census techniques based on standard classification metrics. We examine each passive source seperately and the three sources combined. The four metrics are defined in Section 2.

Table 2 shows that darknets observe a sufficient amount of spoofed traffic that neglecting it would invalidate our inferences. For example, the first heuristic in the table covers approximately 10M /24s covered, whereas our final estimates of active /24 blocks are around 3M per darknet (Section 5). In total, our filters removed over 7.2M /24s from the traffic data at each darknet. To assess the impact of the remaining spoofing, we examined the portion of the remaining filtered traffic that had source addresses we knew to be spoofed because: either (i) they originated from UCSD and MERIT darknet IP addresses, or (ii) from /24 blocks monitored at SWITCH from which we never observed a bidirectional flow (4574 /24 blocks out of the 9343 total /24 blocks monitored at SWITCH). Table 3 summarizes this analysis, showing that our filters captured most traffic using source addresses that we know to be spoofed.[6] The substantial reduction suggests the remaining spoofing is low.

4.3 Validation

Table 4 summarizes the validation results from applying our techniques to data from three sources (UCSD, MERIT, and SWITCH), using the validation data set constructed from BGP and ISI as described in Section 3. We report the precision, recall, true negative rate, and accuracy, for each source separately and in aggregate. We compute these metrics based on a labeled data set that includes 10.8M /24 labeled blocks (4.3M positives and 6.5M negatives). This labeled data has two key limitations. First, though we compiled a large labeled dataset of several million prefixes, the negatives are based on information about *unrouted* networks, which are not representative of *routed but unused* networks in the Internet; this bias can lead to both overestimating or underestimating the accuracy of our techniques. Manual analysis of darknet traffic also revealed a few unrouted address blocks that seem to be used internally (but not globally advertised) by some organizations (see Section 5). Secondly, our positives are based on destinations that respond to ISI Census probes, which may come from from border routers rather than end hosts (see Section 5); these will induce false positives in the ISI data and our labeled data set, which may induce under-estimation of performance of our method.

Table 4 shows the strong performance of our techniques in terms of precision, true negative rate, and accuracy. High precision means that the blocks we infer as used are actually used in most cases. The lower values for recall show that our techniques do not capture all active /24 blocks, which is consistent with the fact that each of our measurement sources sees only a substantial (64.5% - 75.6%) fraction of the labeled positives. Combining measurements from all three sources increases recall to 0.811. Our techniques also yield a high true negative rate, above 0.999, i.e., they correctly identify the vast majority of unrouted networks as unused. Finally, the last column of Table 4 shows that the overall accuracy, including negative and positive samples, is between 0.859 and 0.903 and improves when combining our three data sources to 0.924.

tined to a single IP, with a payload of all z's.

[6]We do not report combinations with source and destination addresses in the same darknet (e.g., UCSD-to-UCSD); our final algorithm excludes packets with sources from known darknets.

	Number of /24 blocks	% of routed address space
UCSD	3,139,366	30.7%
MERIT	2,982,609	29.1%
SWITCH	3,633,905	35.5%
All passive	3,942,605	38.5%
ISI	4,281,875	41.8%
Total	4,753,093	46.4%

Table 5: Number of active /24 blocks discovered by each census method. The methods of estimating address space usage discussed in this paper have a considerable overlap in the /24 blocks they observe. By combining methods we increase the number of /24 blocks known to be active.

5. A FIRST LOOK AT THE IPV4 MAP

A complete evaluation of the differences in results given by our passive approach versus applying data from ISI's active approach to a /24 block granularity is beyond the scope of this paper, but we examine the most obvious differences and try to explain them based on manual analysis.

The Hilbert map in Figure 2 compares the combined results from our three passive measurements (Switch + UCSD + Merit) to the ISI Census data filtered (Section 3) and aggregated at a /24 granularity – a high-resolution version of this image is available at [7]. Table 5 provides the count of /24 blocks discovered by each approach and also reports the union of these sets. We find several large contiguous address blocks, e.g., /8 to /12, that appear largely populated in only one of the two measurement approaches.

Blocks that the ISI Census estimates as mostly used but that our passive traffic approach infers as entirely unused (other than some noise that we assume are filtering errors) are solid green in Figure 2. Manual inspection of the largest

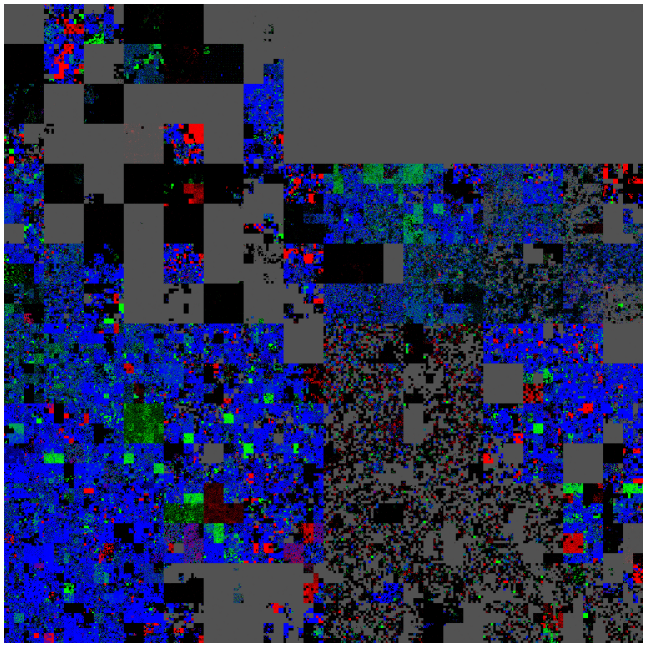

Figure 2: Hilbert map visualization comparing merged passive (UCSD, MERIT and SWITCH) datasets with ISI Internet Census data. The IPv4 address space is rendered in two dimensions using a space-filling continuous fractal Hilbert curve of order 12 [18, 21]. Each pixel in the full-resolution image [7] represents a /24 block; red indicates blocks observed only in the passive data, green blocks are only observed in ISI Census data, and blue blocks are in both. Unrouted networks are grey. The map highlights differences between inferences from passive and active measurements, including significant activity (according to the former) in two /8 legacy allocations.

block (a /8 network) exhibiting this behavior, suggests that most responses observed by ISI are from routers that consistently respond to ICMP echo requests to addresses ending with the same byte (e.g., routers may be configured with virtual interfaces ending in .1 for multiple /24 subnets). Our /24-granularity aggregation of ISI Census data leads us to infer, in such /8 block, nearly 12,000 active /24 blocks (59% of all /24 blocks marked as active) solely due to responses to probes toward addresses ending in .1. Responses to these probes indicate an initial TTL of 255, almost exclusively used by routers and Unix boxes [20]. Similarly, blocks that, when aggregating ISI Census data in /24 blocks, appear as heavily populated, but we inferred as only lightly populated (mostly green with some blue in Figure 2) are also likely due to routers. Another possible interpretation of this discrepancy is that both phenomena occur: (i) networks matching such blocks have a lower fraction of infected machines compared to other networks and/or policies that block outgoing known malicious traffic (e.g., a filter on port TCP 445 would block all Conficker-like traffic), resulting in their under-estimation from darknet measurements; (ii) such networks do not exchange traffic with the SWITCH research network.

Another type of discrepancy we identify is blocks that are solely present in the passive datasets (solid red in Figure 2). Manual inspection of three of these cases reveals two possible causes. If only the darknet, but not the live network, shows these blocks as largely populated, it may be spoofed traffic that our filters did not catch. However, in two of these three cases, we observed traffic from such network blocks in both the dark and live networks; we speculate that the ISI Census may have under-estimated the population, either because they did not probe the network (due to operator request) or because those networks filter ICMP requests.

Perhaps of most interest are the approximately 15 /8 networks that the ISI Census finds to be largely unused. These are mostly legacy allocations to organizations that vastly under-utilize them. Our passive measurement techniques also confirm that the majority of these networks appear as unused, although at least two of them, both legacy allocations, appear largely active. One block is allocated to a large electronics company, and while both the ISI Census and darknet datasets show limited use of the block (83 and 245 /24 blocks respectively), the SWITCH dataset reflects much higher use (942 /24 blocks). This difference may mean that the organization (i) filters ICMP requests, thus limiting its visibility by the ISI Census, and (ii) has little darknet-observable malware on internal hosts, or filters it outbound, limiting its visibility by darknets. The other /8 network is allocated to a large communications provider which also likely filters ICMP requests, reducing its visibility via active measurements. In this case, 99% of the networks identified as active exclusively by the passive technique are found in a single /10. Of this /10, the passive approach identifies 6,777 /24 blocks as active, while none are active according to the ISI Census dataset.

We also verified that the portions of both /8 networks that our traffic measurements show are active are mostly not considered active even in the latest (2013) ISI Census [14]. These examples illustrate that a passive traffic-based estimation can reveal insight into IPv4 address utilization beyond what can be seen with active measurements.

Finally, our passive measurements show active blocks in a few unrouted networks (shaded in grey in Figure 2). We expected all unspoofed traffic coming from unrouted space to use private IP addresses. However, in our darknet measurements we found about 140 unrouted, non-private networks sending traffic that is unlikely spoofed. One of the largest classes of traffic seen in general by the UCSD network telescope is "conficker-like," or TCP packets to port 445. Such traffic has some consistent idiosyncracies: (i) a bug in the pseudorandom number generator that causes them to send packets to only a quarter of the address space [8]; (ii) TCP SYN retransmits in attempts to open a connection, with specific inter departure times [4]. We observed exactly these traffic patterns from these unrouted blocks. This traffic is not visible in the SWITCH measurements because is filtered out by our live-network filtering heuristic. Such unspoofed traffic from unrouted space could result from organizations using their assigned space privately without globally advertising it, or privately using IP addresses not assigned to them as if they were RFC1918 addresses [1].

6. RELATED WORK

Heidemann *et al.* [10] was the first published census of IPv4 address activity using active network probing; they minimized probing overhead and associated complaints by spreading the scan over a 30+-day window. Zander *et al.* [26] estimated the *number* of IPv4 addresses actually used on the Internet by combining active probing and additional data such as IP addresses in Wikipedia logs, spam blacklists, web server logs. They estimated approximately 1 billion IPv4 addresses used, which is around 40% of the publicly routed space. The popularity of active probing methods motivated the deployment of several efficient scanning tools, including *zmap* [9], which uses ICMP to scan the entire IPv4 address space in under 45 minutes using a single machine. An anonymous individual (or group) recently published the results of a series of illegal (botnet-orchestrated) scans of the IPv4 Internet address space [11] from over 400 thousand bots. Their ICMP scanning results resemble ISI's Census findings, since they found 4.3M /24 blocks containing about 420M "pingable IPs". They elicited responses from more (36M) IP addresses when scanning for open TCP/UDP ports. However, their probing methodology is not well-documented, and their measurements may be skewed due to local transparent proxies intercepting probes from the bots [2]. They also used the botnet to perform reverse DNS lookups as an alternative method to infer address space usage; their inferences match ours for two legacy /8 allocations not visible by the ISI Census (see Section 5).

We developed a methodology that relies only on passive measurements to infer macroscopic network activity; passive techniques avoid some methodological issues with active probing, most notably policies that prevent responses to active probes or induce responses from addresses other than those probed. However, inferences based on passive data also have limitations, most notably the presence of traffic using spoofed source IP addresses, especially in traffic data that is easiest to obtain, i.e., from IP darknets. The most important contribution in this work is our heuristics to filter out spoofed IP traffic from both darknet as well as operational network traffic flow data (Section 4).

Spoofing is a persistent threat [5, 19, 25], and at least two other studies have used TTL-based inference with active and passive measurements to detect spoofed packets [6,23]. Both approaches try to establish reference values of TTL for different traffic classes, inferring that packets with diverging values are spoofed. In [23], the authors observe TTL values of distinct source IP/protocol pairs over time, to learn which values legitimate hosts use. Beverly [6] developed a supervised learning classification algorithm that considers the TTL value as an indicator of the length of the path an IP packet traverses. His algorithm classifies incoming packets into legitimate IP addresses based on valid origin-destination path lengths (i.e., TTL value). Neither of these methods work for darknet traffic, since we cannot identify a reference TTL value for /24 blocks originating a limited amount of traffic.

7. CONCLUSION

We developed and evaluated a methodology for removing spoofed traffic from data sets collected on both darknets and live networks, and found the resulting filtered data to effectively support census-like analyses of IP address space utilization. Although some spoofed traffic required manual removal, we also identified several general classes of spoofed traffic that enabled us to create heuristic filters to remove them. Passive traffic data allowed us to identify $\approx 450K$ /24 IPv4 blocks as active that were not inferred as active by ISI's most recent census measurements; visibility into other parts of the IPv4 address space would expand with additional vantage points.

One possible future direction of this work is a hybrid approach that first infers active IP address blocks based on passive measurements from one or more (live or dark) vantage points, then probes only addresses that cannot be confidently inferred as active. Using all three passive datasets we gathered, a hybrid approach would not only yield additional discovery of active /24 blocks, it would also reduce the active probing using ISI's method by 38.5%. Using only the SWITCH traffic vantage point with a hybrid approach would increase the inferred active /24 blocks and reduce measurement by $\approx 400K$ and 35.5%, respectively. In this hybrid approach, the marginal utility of adding measurements from the darknets thus seems limited ($\approx 75K$ additional active /24 blocks), but a passive-only measurement scenario would benefit significantly from these additional vantage points ($\approx 300K$ more active /24 blocks than when using only SWITCH data).

This preliminary investigation inspires many additional questions on the strengths and limitations of this methodology. How much does the vantage point matter, in terms of location and size of address space observed? Would traffic measurements from IXPs provide considerably more insight over a shorter time period? Can we improve our ability to detect (and validate) spoofed traffic, perhaps by responding to darknet traffic? For a given segment of address space, do traffic characteristics correlate with present or future address utilization patterns? How well will this technique work for IPv6? We hope our results encourage others to investigate the potential to exploit passive Internet traffic measurements to perform Internet-wide census studies.

Acknowledgements

UCSD network telescope operations and data collection, curation, analysis, and sharing is provided by NSF CRI CNS-1059439 and CNS-1228994, DHS S&T NBCHC070133 and UCSD.

8. REFERENCES

[1] `http://seclists.org/nanog/2009/Feb/2`.
[2] A. Dainotti, A. King. CAIDA Blog: Carna botnet scans confirmed. `http://blog.caida.org/best_available_data/2013/05/13/carna-botnet-scans/`.
[3] Advanced Network Technology Center, University of Oregon. Route Views Project. `http://www.routeviews.org/`.
[4] K. Benson, A. Dainotti, k. claffy, and E. Aben. Gaining Insight into AS-level Outages through Analysis of Internet Background Radiation. In *Traffic Monitoring and Analysis Workshop (TMA)*, Apr 2013.
[5] R. Beverly and S. Bauer. The spoofer project: inferring the extent of source address filtering on the internet. In *USENIX Workshop on Steps to Reducing Unwanted Traffic on the Internet*, SRUTI'05.
[6] R. E. Beverly, IV. *Statistical learning in network architecture.* PhD thesis, MIT, 2008. AAI0820515.
[7] CAIDA. Supplemental data: Estimating Internet address space usage through passive measurements.

http://www.caida.org/publications/papers/2013/passive_ip_space_usage_estimation/supplemental/, 2013.

[8] E. Chien. Downadup: Attempts at Smart Network Scanning. http://www.symantec.com/connect/blogs/downadup-attempts-smart-network-scanning, 2009.

[9] Z. Durumeric, E. Wustrow, and J. A. Halderman. ZMap: Fast Internet-wide scanning and its security applications. In *Proceedings of the 22nd USENIX Security Symposium*, 2013.

[10] J. Heidemann, Y. Pradkin, R. Govindan, C. Papadopoulos, G. Bartlett, and J. Bannister. Census and survey of the visible Internet. In *8th ACM SIGCOMM conference on Internet measurement*, IMC '08.

[11] J. Horchert and C. Stöcker. Mapping the internet: A hacker's secret internet census. Spiegel Online, March 2013.

[12] Information of Sciences Institute, University of Southern California. LANDER project:Internet address census it49c-20120731. http://www.isi.edu/ant/traces/internet_address_census_it49c-20120731.README.txt, 2012.

[13] Information of Sciences Institute, USC. Internet Address Survey Binary Format. http://www.isi.edu/ant/traces/topology/address_surveys/binformat_description.html, 2012.

[14] Information of Sciences Institute, USC. ANT Census of the Internet Address Space - browsable map. http://www.isi.edu/ant/address/browse/index.html, 2013.

[15] Insecure.Com LLC. Nmap Security Scanner. http://nmap.org.

[16] A. Langley. Probing the viability of TCP extensions. Technical report, Google Inc., Sep 2008.

[17] Merit Network, Inc. Merit Darknet IPv4. http://software.merit.edu/darknet/.

[18] R. Munroe. xkcd: MAP of the INTERNET 2006. http://blog.xkcd.com/2006/12/11/the-map-of-the-internet/.

[19] 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*, IMC '04, pages 27–40, New York, NY, USA, 2004. ACM.

[20] A. Sebastian. Default Time To Live (TTL) values. http://www.binbert.com/blog/2009/12/default-time-to-live-ttl-values/, 2009.

[21] A. N. Shannon V. Spires. Exhaustive search system and method using space-filling curves. Patent, 10 2003. US 6636847.

[22] SWITCH. Swiss Tele Communication System for Higher Education. http://www.switch.ch/.

[23] S. Templeton and K. Levitt. Detecting spoofed packets. In *DARPA Information Survivability Conference and Exposition*, 2003.

[24] University of California, San Diego. The UCSD Network Telescope. http://www.caida.org/projects/network_telescope/.

[25] E. Wustrow, M. Karir, M. Bailey, F. Jahanian, and G. Huston. Internet background radiation revisited. In *10th ACM SIGCOMM conference on Internet measurement*, IMC '10, 2010.

[26] S. Zander, L. L. H. Andrew, G. Armitagei, and G. Huston. Estimating IPv4 Address Space Usage with Capture-recapture. In *IEEE Workshop on Network Measurements (WNM 2013)*.