Lost in Space: Improving Inference of IPv4 Address Space Utilization

Alberto Dainotti, Karyn Benson, Alistair King, Bradley Huffaker, Eduard Glatz, Xenofontas Dimitropoulos, Philipp Richter, Alessandro Finamore, Alex C. Snoeren

alberto@caida.org
MAPPING USE OF IPV4 SPACE

Why do we care?

Security
- inform host reputation and access control
  - e.g., to mitigate network abuse
- detect BGP hijacking attacks

Policy / Social / Economics
- inform policy on address space allocation
- estimate Internet usage over time
  - e.g., policy, political and social science, economics

Better Data Analysis
- identify homogeneous address aggregates
  - e.g., for IP geolocation
- data normalization
  - e.g., per-AS or per-Country normalization

Operations
- detect changes in network operation
- select targets for active measurements
  - e.g., traceroutes, vulnerability scans, and reachability surveys

Center for Applied Internet Data Analysis
University of California San Diego
- active + passive measurement approaches
- passive: main challenge is filtering out *spoofed* and *scanning* traffic
A TAXONOMY OF SPACE USE

announced on BGP does NOT imply it is used
Center for Applied Internet Data Analysis
University of California San Diego

A PEEK AT RESULTS

Paper/website are packed with graphs and tables

Our filtering yields 10.4M routed /24 blocks that we further classify as used or unused. We define a /24 block as used if at least one of its IP addresses is assigned to a machine that will exchange packets on the public Internet with such address in the IP header. In Sections V and VI, we discuss the inference methodologies—based on both active and passive measurements—that we use for this purpose. Figure 2 provides an overview of our final results according to our taxonomy and breaking the space by RIR and legacy allocations. This visualization succinctly represents "where" in the allocation system, and how, large portions of address space appear unutilized.

V. ANALYSIS OF PASSIVE TRAFFIC

Is the approach of passive measurement for inferring address space utilization generally applicable? How does it depend on different network types, trace types, and other parameters? We first extend the method of [20], which used data from a darknet and an academic ISP, to work with the fundamentally different types of traffic collected at a residential ISP and an IXP, showing how to filter out spoofed traffic in different trace types (Section V-A). Second, we evaluate the impact on our inferences of varying aspects of the vantage points: traffic composition, size of monitored address space, duration and time of measurement (Section V-B).

September 17, 2015
DRAFT
Fig. 8: Hilbert map visualization showing the utilization of the address space according to our taxonomy. The IPv4 address space is rendered in two dimensions using a space-filling continuous fractal Hilbert curve of order 12 [53], [58]. Each pixel in the full-resolution image [17] represents a /24 block; light blue indicates used, dark blue routed unused, purple unrouted assigned, black unassigned, and grey reserved by RFC blocks.

(5.1M /24s) to the entire used address space (5.3M /24s)! ARIN, RIPE, APNIC, and LACNIC have 50%, 65%, 54% and 68% of their address blocks used, respectively, in contrast to AFRINIC which has fewer of their blocks used (31%) and many more available (38%) address blocks than other RIRs (6.7% of other RIR addresses are available).

Table IX(a) lists the top-5 continents and countries in routed unused and unrouted assigned /24s. 52.2% of the routed unused space and 72% of unrouted assigned space is in North America, primarily in the U.S., where most legacy allocations were made. Asia follows, with China owning 8.79% and 5.7% of the global routed unused and unrouted assigned space, respectively, and then Europe. Other continents (South America, Oceania, and Africa) have between 0.93% and 2.13% of the global routed unused and unrouted assigned space. Figure 9 visually illustrates the per-country ratio of assigned unused (the sum of routed unused and unrouted assigned) over assigned (that is, usable minus available) space, suggesting which regions are using space most and least efficiently. The U.S. is red in this map due to a few very large allocations, while some African countries are red because they use a very small fraction of their (also small) assigned space.

Figure 10 compares address space assigned to countries to per-country population [19] and Gross Domestic Product (GDP - we used "purchasing power parity" from CIA’s World Factbook [18]). We observe notable disparities between used /24s and population. For example, September 17, 2015.
CONTRIBUTE

how you can help us

- data data data...
  - anonymized, no timestamps, no content
  - just /24 blocks you observe (NOT IPs of your network) within a large time frame (e.g., 3 months or even a year)
  - 1) existing: server logs, NetFlow records…
  - 2) collect: host a simple box running a modified Tstat. opensource code (we don’t need access to the machine)

- we share our resulting dataset through the DHS’ Protected Repository for the Defense of Infrastructure Against Cyber Threats (PREDICT) www.predict.org
THANKS

questions?

alberto@caida.org