Errata for SIGCOMM CCR (Vol. 44, Issue 1, January, 2014) 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

This errata is to help viewers/readers identify/properly understand our contribution to the SIGCOMM CCR Newsletter. Volume 44 Issue 1, (January 2014) on pages 42-49.

<http://dl.acm.org/citation.cfm?doid=2567561.2567568>

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

We discovered three errors in our analysis for this paper, which have slight effects on the results we reported in Sections 1, 4, and 5.

The first error was a syntax error in the filter we used to exclude spoofed source traffic; this filter erroneously excluded 170,501 and 153,974 /24 blocks respectively from the UCSD and MERIT darknet traffic samples, although some of these we observed at the SWITCH vantage point anyway. Once we corrected this error, the net total number of /24 blocks discovered by our passive methods increased by 65,551 (from 3,881,054 to 3,942,605), and the number of such blocks that were not also observed via our active methods increased by 19,211 (from 452,007 to 471,218). We had also reported that our passive measurements for conducting an Internet census could reduce the overhead of an active approach by 37.9%, but correcting our filters increased this overhead reduction to 38.5%.

4. TECHNIQUES FOR A PASSIVE CENSUS

4.2 Measurements from darknets

The second error we discovered is that we mislabeled some of the SWITCH /24 blocks when marking as “dark” those from which we never observed a bidirectional flow. Correcting this mislabeling yielded 4574 /24 dark blocks in SWITCH rather than the previously reported 5224. After

Monitored destination			Number of /24 blocks (sources)	
			MERIT	SWITCH-DARK
UCSD		before filtering	54210 (98.4%)	4522 (98.9%)
		after filtering	21 (0.038%)	0 (0%)
MERIT		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). Before filtering, these blocks appear 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 separately and the three sources combined. The four metrics are defined in Section 2 of [2].

fixing both the labeling of SWITCH blocks and the filter to exclude spoofed traffic, we obtained slightly better results, shown in (an updated) Table 3.

4.3 Validation

The third error we discovered was a miscalculation in recall and accuracy when evaluating the combined results from the three vantage points. We updated Table 4 to reflect such miscalculations: recall and accuracy for the “Total” category are 0.811 and 0.924 respectively instead of the previously reported 0.906 and 0.962. Table 4 also shows new values for UCSD and MERIT, obtained considering the previously erroneously excluded traffic. In particular, the recall and accuracy improved for both darknets (UCSD: recall improved from 0.641 to 0.672 and accuracy improved from 0.857 to 0.869; MERIT: recall improved from 0.615 to 0.645 and accuracy improved from 0.847 to 0.859).

5. A FIRST LOOK AT THE IPV4 MAP

After correcting these errors, we updated the Hilbert map (new Figure 2 below), the totals observed by each darknet and by aggregating sources (Table 5), and the specifics of the noted discrepancies between active and passive measurements in legacy /8 allocations. There are no striking differences to the updated Hilbert map. Both darknets observe 1.5% to 1.7% more of the total routed address space (MERIT increases its visibility from 27.6% to 29.1% of the total routed address space while UCSD increases from 29.0% to 30.7%). When considering all passive vantage points, the increase was less significant because we observed many of these /24 blocks at the SWITCH vantage point anyway. Similarly, the increase in /24 blocks observed by combining the passive and active measurements was only 0.1% of the total routed address space. In the legacy /8 allocations, after applying our fixes, the darknet observed 91 additional /24 blocks belonging to a large electronics company, and 127 blocks belonging a large communications provider.

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 [3, 4]. Each pixel in the full-resolution image [1] 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. (caption unchanged from original paper, though heatmap is slightly different).

	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.

7. CONCLUSION

In the conclusion of the original paper, we discussed a hybrid approach where the blocks inferred via passive measurements as used could be excluded from active probing. The updated list of /24 blocks captured in the UCSD and MERIT darknets implies that:

- We could reduce active probing requirements by 38.5% (instead of the previously reported 37.9%).
- The marginal utility of adding the darknet measurements to the SWITCH data in the hybrid approach was $\approx 75K$ /24 blocks (instead of the previously reported $\approx 50K$ /24 blocks).
- The darknets observed $\approx 300K$ /24 blocks that were not observed at SWITCH (instead of the previously reported $\approx 250K$).

7. REFERENCES

- [1] 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.
- [2] A. Dainotti, K. Benson, A. King, k. claffy, M. Kallitsis, E. Glatz, and X. Dimitropoulos. Estimating Internet address space usage through passive measurements. In *ACM SIGCOMM Computer Communication Review (CCR)*, 2014.
- [3] R. Munroe. xkcd: MAP of the INTERNET 2006. <http://blog.xkcd.com/2006/12/11/the-map-of-the-internet/>.
- [4] A. N. Shannon V. Spires. Exhaustive search system and method using space-filling curves. Patent, 10 2003. US 6636847.