Exploring Adaptability in Smart Internet Scanners

Kevin A. Negy, BS in progress[1], Austin Murdock, PhD in progress[2], Frank Li, PhD in progress[2] and Vern Paxson, PhD[2]

[1] Computer Science, University of Central Florida, Orlando, FL, [2] Electrical Engineering and Computer Sciences, University of California, Berkeley, Berkeley, CA

Introduction

Internet-wide network scanning allows researchers to analyze the status of connected hosts and measure Internet trends. Such scanning techniques have been developed and fruitfully used for the current protocol (IPv4) used to assign Internet addresses to devices. However, the Internet is moving increasingly towards IPv6, where existing scanning techniques are not feasible. In particular, the address space, or set of addresses, used in the current version is small enough ($2^{32} \approx 4.3 \times 10^9$) to exhaustively scan, however the address space used in the new protocol is not ($2^{128} \approx 3.4 \times 10^{38}$), being exponentially larger. Smart scanners must now use intelligent methods to select and scan regions of the space likely to contain many hosts.

Background

Target Generation Algorithms (TGAs) are a specific type of internet scanning algorithms that being with a group of **seeds**, or a pre-determined set of known addresses, that are used to generate a list of **target addresses.** These target addresses are scanned, their responses are recorded, and the effectiveness of the TGA can be determined. Previous types of TGAs are:

-Pattern-based scanning[1] – creates target addresses based on seed patterns
-Entropy/IP[2] – uses machine learning and Bayesian networks to determine statistical dependencies of the seeds
-6Gen[3] – uses seed clusters to create *n* address **ranges** to scan, where $\sum_{i=1}^{n} size(i) \leq$ **probe budget** (user defined input)

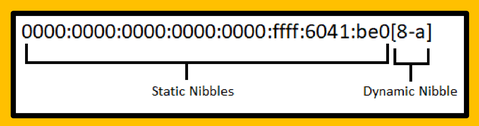


Figure 1: Example IPv6 address range. Range size = 3.

Objective

In this project, we use empirical measures to explore the improvements of **adaptive scanning** building off of 6Gen. Due to time constraints, 6Gen selects regions of the address space to probe based on a given set of seeds and then probes all selected regions to completion with no regard for response rate. Adaptive scanning allows for dynamic modifications to region size during the scan, i.e. low response rates during run-time would terminate a region scan and high response rates would cause the system to expand the region.

Method

Adaptive Algorithm
1) Choose ranges with
   -A size of more than 1000
   -Hit rate of more than 5%
   -Hit rate of less than 95% (to account for aliasing)
2) Assign equal number of probes for each range up to probe budget
3) Expand regions without going over budget
   -From right to left, expand dynamic nibbles
   -From right to left, expand static nibbles

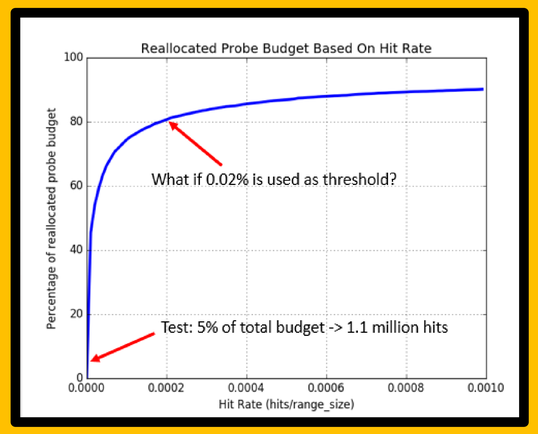
Example: Probe budget: 80
2000::008[3 - 4]
=> 2000::00[6 - a][0 - 3] = 5*4 = 20
=> 2000::00[6 - a][4 – f] = 5*12 = 60
Total target addresses = 80

Results

We used a probe budget of 100 million. The adaptive scanner produced a set of ranges whose overall size was as close to the probe budget as possible. Then, we removed duplicate, blacklisted, and previously scanned addresses. This left 94.8 million target addresses to scan.

	Pre-adaption ranges	Adapted ranges
Budget	-	100 million
Target addresses	1,783 million	94.8 million
Hits	33.4 million	1.1 million
Hit Rate	1.87%	1.17%

Although, the overall hit rate appears to decrease when comparing only-adapted ranges to the original ranges, the adaptive ranges only needed to outperform the low-response ranges whose probes were reallocated. This is successful, as the test budget of 100 million would have been obtained from large ranges of close to 0% hit rate. Therefore, with only 5% of the budget reallocated, we gain 1.1 million hits. We potentially could gain several million if we had reallocated 80% of the budget while only having to beat a percentage of .02% for adaption to be worthwhile.



Conclusion

Incorporating adaptive scanning to 6Gen is a promising next step to improving overall IPv6 hits. Full implementation could yield millions of new hits at no extra probe cost.

Future Work:
- Incorporate adaptive, decision-making rounds
- Distribute reallocated budgets unequally to higher rate ranges
- Incorporate pattern recognition techniques

References

1. J. Ullrich, P. Kies Murdock et al., ACM Internet Measurement Conference 2017 eberg, K. Krombholz, and E. Weippl. On Reconnaissance with IPv6: A Pattern-Based Scanning Approach. In Availability, Reliability and Security (ARES), 2015 10th International Conference on, pages 186–192. IEEE.

2. P. Foremski, D. Plonka, and A. Berger. Entropy/ip: Uncovering structure in ipv6 addresses. In Proceedings of the 2016 ACM on Internet Measurement Conference, pages 167–181. ACM, 2016.

3. Murdock et al., 6Gen: Target Generation for Internet-wide IPv6 Scanning., In Proceedings of the 2017 ACM on Internet Measurement Conference (to appear).

Acknowledgements

This research was funded by the SUPERB – ITS REU through the National Science Foundation, Grant No. 1659833.