Chair of Network Architectures and Services
Department of Informatics
Technical University of Munich

Thesis B.Sc.

Thesis M.Sc.

Maximize Information Gain from Internet Scans while Minimizing Impact

Motivation

Active scanning can be a a method to gain information about the Internet. A scan is usually done by a sending a single probe to a massive amount of hosts. But the responses always depend on used probe and the amount of information gained is always a subset of what is possible to collect. How can we gain more information?

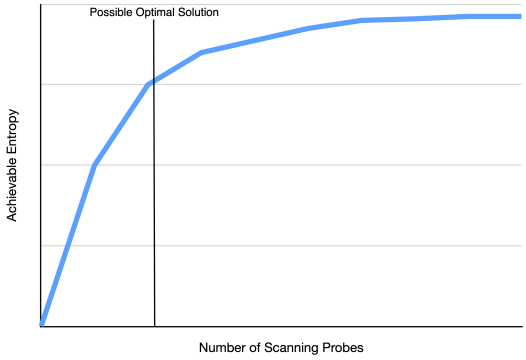

Figure 2: The most entropy achievable by a certain number of probes

Approach

A solution would be to define multiple probes. But what probes and how many are needed? If too many are used a scan would take too long and the amount of request are comparable to a DoS attack. If too few or the wrong ones are used the information obtained is too low.
Formal definitions for information and entropy can be used to model the problem. Afterwards we can develop an algorithm to find the optimal set of probes to maximize the entropy we gain from each host. Such an algorithm is not trivial as already with 1000 probes, a total pf 2^{1000} possibilities exist to combine them (too much to compute). So brute force is not an option and the findings from the formal modeling should enable us to develop a better algorithm.
Most likely their is not a single solution, but multiple ones. With a multi-objective (pareto) optimization a curve like fig. 2 can be constructed. Afterwards an optimal solution can be picked that is the best trade-off between information and impact on other servers (= number of probes).

Your Task

- Formal modeling of the problem with help of the information theory
- Develop and implement an algorithm that finds the optimal set of scanning probes for maximum entropy
- Apply the results on Internet scans and evaluate the scan data

Contact

Markus Sosnowski sosnowski@net.in.tum.de
Patrick Sattler sattler@net.in.tum.de
http://go.tum.de/720506