

Development of an Efficient Large Scale DNS Scanning Pipeline

Motivation

The DNS builds the foundation to all robust and modern communication systems over the Internet. The chair performs several Internet wide measurement campaigns and the basis to many of these are DNS measurements. We regularly collect domain names from different sources and resolve these. We currently use MassDNS [1] pointed towards a local deployed unbound resolver [2] to resolve our input lists. The current process to resolve all lists for all needed resource records is hard to maintain and does not provide a simple extensibility approach. As this data is used by other scanning processes it is important that the resolution is reliable and robust. All detected errors should be reported in a way that the maintainer of the system can understand the issue and act quickly.

Another new requirement for the pipeline is a daily scan of important input lists. Currently, our ethical scanning limits do not allow to resolve all domains within 24 hours. Nevertheless, we need some lists (e.g., top lists like the Alexa Top 1 Million [3]) to be scanned once a day. Moreover, in order to also support large scale data analysis of the results, the resulting format needs to support state of the art analysis tools. Therefore, we want to develop a new pipeline which also fulfills all these new requirements.

Your Task

- Familiarize yourself with the main used tools (MassDNS, Unbound...)
- Understand the requirements and their accompanying problems
- Develop a pipeline with a focus on
 - Reliability
 - Robustness
 - Extensibility
 - Maintainability

Requirements

- Basic programming knowledge in the Shell/Bash scripting language
- Familiarity with GIYF-Based work approaches
- Good code quality in order for others to continue and understand your work
- Optionally: Knowledge on how to write makefiles

Bibliography

- [1] <https://github.com/blechschmidt/massdns>
[2] <https://nlnetlabs.nl/projects/unbound/about/>
[3] <https://www.alexa.com/topsites/>

Contact

Patrick Sattler sattler@net.in.tum.de
Johannes Zirngibl zirngibl@net.in.tum.de

<http://go.mytum.de/286232>

