Project Proposal

Predicting the Origin of DNS Traffic Without Reference to Client Source Address

Joe Abley

September 28, 2018

1 Problem Description

The Domain Name System (DNS) includes a wire protocol with which structured requests and responses are exchanged over a network. The DNS protocol is specified and widely used using the UDP transport protocol. Since UDP is stateless there is no way for the receiver of a DNS request sent over UDP to verify the legitimacy of a source address. A consequence of this is that DNS servers are frequently used as amplifiers in reflection attacks [2]. Although some such attacks are trivially identified, for example by Query Type (QTYPE), many are more difficult. By choosing query parameters that match legitimate, real-world use of the DNS, the attacks may seem impractical to block without collateral damage. This is especially true of amplification attacks against DNS resolvers.

The goal of this project is to construct a classifier that can distinguish between a stream of DNS queries received from one real-world client and a stream received from another without reference to source address, based on training sets derived from queries whose origin is accurately known.

2 Data Sets

Afilias operates authoritative DNS servers for the INFO top-level domain, as well as many others. These servers are made available to the Internet through commodity transit arrangements from multiple locations using anycast [1], as well as through so-called Private Network Interconnects (PNIs). The origin of queries received over PNIs can be said to be known with high accuracy; the origin of queries received over the Internet, however, cannot in general be trusted.

Complete sets of query data received over both trusted and untrusted paths are available from Afilias for use in this project in the form of packet captures in PCAP format[1].

Time-series data sets are constructed from query data by identifying features that distinguish different queries and creating a vector of metrics for each class of queries along a time axis calculated within regular time intervals. Examples of features are the proportion of queries with `QTYPE=MX` or the number of labels in the Query Name; corresponding coordinates in the time-series vector might be the mean, standard deviation and 90th percentile of each such metric. This method has been used successfully by Castro and Qiao to build similar classifiers based on DNS query data [4].

The identification of a set of features that allows a useful classifier to be trained is the primary desired result of this project. Choice of algorithm, hyperparameters, etc is expected to be determined automatically using a suitable machine learning toolkit such as `auto-sklearn` [3].

3 Real-World Applications

Traffic received over the Internet that purports to be from a particular client but which is classified otherwise can be isolated itself be subject to classification. Being able to classify known-bad traffic from packet captures after the fact is useful in forensic analysis of strange traffic patterns (e.g. noticeable spikes in traffic volume) and potentially also for use in real-time if the matching criteria can be distilled down into something with low cost to execute, like a Bloom filter.

Finding traffic from known clients arriving from an unexpected direction also has the potential to inform operational management systems, e.g. causing an alarm in the network operations centre that could prompt an escalation. An example of such a client might be Google Public DNS, who are relied upon by a large number of end-users.

References

[1] J. Abley and K. Lindqvist. Operation of anycast services. BCP 126, RFC Editor, December 2006. URL: http://www.rfc-editor.org/rfc/rfc4786.txt.

[2] J. Damas and F. Neves. Preventing use of recursive nameservers in reflector attacks. BCP 140, RFC Editor, October 2008. URL: http://www.rfc-editor.org/rfc/rfc5358.txt.

[3] M. Feurer, A. Klein, K. Eggensperger, J. Springenberg, M. Blum, and F. Hutter. Efficient and robust automated machine learning. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems 28*, pages 2962–2970. Curran Associates, Inc., 2015. URL: http://papers.nips.cc/paper/5872-efficient-and-robust-automated-machine-learning.pdf.

[4] J. Qiao. Source address classification - feature engineering. Accessed: 2018-09-16. URL: https://blog.nzrs.net.nz/source-address-clustering-feature-engineering/.

[1] PCAP, named after the C library `libpcap`, is the file format used by the `tcpdump` utility.