Demo Abstract: A Tool to Detect and Visualize Malicious DNS Queries for Enterprise Networks

Jawad Ahmed*,†, Hassan Habibi Gharakheili*, Qasim Raza*, Craig Russell†, and Vijay Sivaraman*
*Electrical Engineering and Telecommunications, University of New South Wales, Sydney, Australia.
†CSIRO Data61, Sydney, Australia.
Emails: {j.ahmed, h.habibi}@unsw.edu.au, q.raza@student.unsw.edu.au, craig.russell@data61.csiro.au, vijay@unsw.edu.au

***Abstract*—This demo presents our web-tool to access and visualize real-time detection of malicious DNS queries for an enterprise network of a large university campus in Sydney, Australia. We showcase two aspects: (1) how to access and process our open data-set containing more than one million DNS queries pertaining to data exfiltration we generated in our campus network, enabling insights into the attributes of such malicious queries; and (2) visualizing our real-time learning-based detection engine operational on 10 Gbps traffic streams from the network border of the university campus.**

I. Introduction

Enterprise networks constantly face the threat of valuable and sensitive data being stolen by cyber-attackers. Sophisticated attackers are increasingly exploiting the Domain Name System (DNS) channel for exfiltrating data as well as maintaining tunneled command and control communications for malware. This is because DNS traffic is usually allowed to pass through enterprise firewalls without deep inspection or state maintenance, thereby providing a covert channel for attackers to encode low volumes of data without fear of detection. The resulting damages can be huge, amounting to several million dollars in a single attack [1]. Several high-profile data exfiltration breaches have been reported recently, for example the Sally Beauty breach (a theft of 25K credit cards) [2].

One way for the attacker to exploit DNS is to register a domain (*e.g.,* `foo.com`) so that the attacker's malware in a host victim can then encode valuable private information (*e.g.,* credit card numbers, or login passwords) into a DNS request of the form `info.foo.com`. This DNS request gets forwarded to the authoritative server for the `foo.com` domain (under the attacker's control).

We [3] develop, tune, and train a machine learning algorithm for real-time detection of anomalies (*i.e.,* exfiltration and tunneling) in DNS queries using a known dataset of benign domains as ground truth. This demonstration complements our efforts outlined in [3]. The contributions of demo are of two-fold: (1) we show how to access our open data-set containing more than a million malicious DNS queries generated using an exfiltration tool [4]; and (2) we demonstrate our tool to visualize the detection of malicious out-going DNS queries in real-time allowing the user to select queries for a specific primary domain, or choose a threshold of the anomaly score given by our machine.

II. Malicious DNS Queries Data

For ground-truth malicious instances, we have generated DNS exfiltration queries by our open source tool, forked from an open source project called "DNS Exfiltration Toolkit" (DET) [4]. We ran our tool on a machine inside the University network that exfiltrates the content of a CSV file containing 1000 samples of random credit card details (obtained from [5]) to an authoritative name server under our control located in a Research network. DET employs AES-256 encryption and uses two tuning parameters namely max length of query name (*i.e.,* 50 to 218 characters) and max length of labels (*i.e.,* 30 to 63 characters) to diversify our synthetic malicious queries. We generated a total of 1.4M exfiltration queries which are publicly available at [6] in form of a CSV file.

Table I lists samples of benign and malicious query names with "unusual" length and string pattern. For example, the malicious query names at the top of this list contains 136 characters. We note that the sub-domain portion of the query names comprises random-looking strings with a significant number of upper-case and numerical characters, and is fairly long. For example, the first malicious query name from the top (*i.e.,* for `cspg.pw`) contains 38 numeric characters (*i.e.,* 28%), and the second malicious query name (*i.e.,* for `29a.de`) contains 23 uppercase letters (*i.e.,* 39%). Given these observations, we define [3] our attributes by three main categories namely characters count, entropy (an indication of randomness) of string, and length of discrete labels in the query name.

Each record of the data represents a DNS query along with its attributes (used as inputs to our anomaly detection model) comprising *query length* (*i.e.,* total count of characters in the fully qualified domain name - FQDN), count of characters in *sub-domain*, count of *uppercase* characters and count of *numerical* characters in the query name, *entropy* of the FQDN, *number of labels*, *maximum label length*, and *average label length* – labels are separated by dots in DNS query names. We encourage other researchers to use and analyze our open dataset drawing interesting insights into attributes of malicious DNS queries.

III. Visualizing Detection of Malicious DNS Queries

We developed a tool to provide an intuitive user-interface for real-time monitoring of outgoing DNS queries in enterprise networks, using ReactJS. Our web-tool is publicly available at [7].

978-3-903176-15-7 © 2019 IFIP

TABLE I
A SAMPLE LIST OF MALICIOUS AND NORMAL DNS QUERIES WITH UNUSUAL LENGTH.

Query name (FQDN)	Security
708001701462b7fae70d0a28432920436f70797269676874.20313938352d32303031204d696372.6f736f667420436f72702e0d0a0d0a0.433a5c54454d503e.cspg.pw	Malicious
PzMnPiosOD4nOCwuOzomPS4nNjovPS8uOzsnNCstODkjOCwoMwAA.29a.de	Malicious
p4-ces3lawazdkbw-qlrq5qalxdt7tycq-385202-i1-v6exp3.ds.metric.gstatic.com	Normal
_ldap._tcp.AWS._sites.dc._msdcs.AD.us-east-1.ec2-utilities.amazonaws.com	Normal

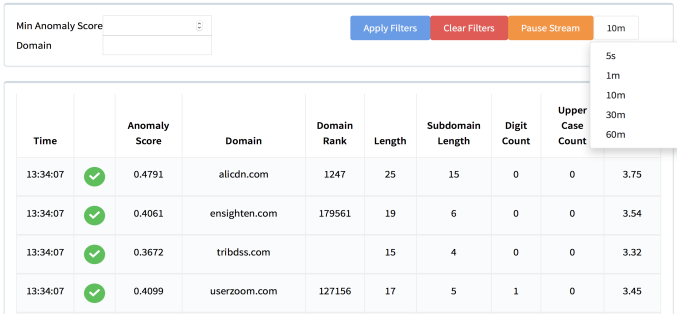

Fig. 1. Web-UI of our real-time DNS exfiltration and tunneling detector.

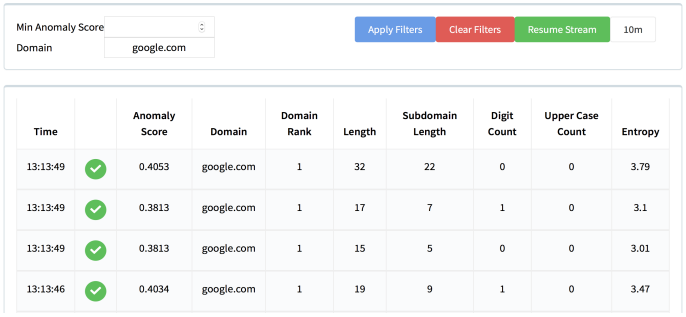

Fig. 2. Filtering queries for a specific primary domain name.

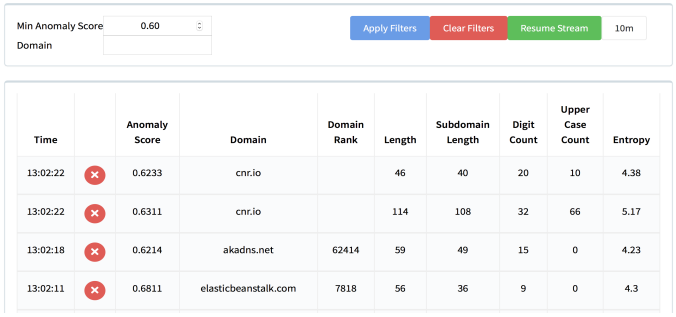

Fig. 3. Filtering queries with a minimum anomaly score.

This tool allows the user to visualize timestamped queries in real-time at various time scales including last 5 sec, 1 min, 10 min, 30 min and 60 min, as shown in Fig. 1. It also shows selected attributes for each query and the output of our model. Records marked by a red cross correspond to DNS queries that are detected as malicious by our machine learning model while benign queries are marked by green ticks – for privacy reasons we do not show the FQDNs of live traffic in our web-tool. The "'domain rank" shows the reputation of the primary domain obtained from Majestic list [3] – if a primary domain is not found in top million Majestic list then its rank is shown empty, for example `tribdss.com` in Fig. 1.

The user can enter a specific primary domain name, filtering all records (in real-time) associated with that primary domain. For example, in Fig. 2 the user has selected `google.com`, the tool is showing the results specific to this filtered domain. Furthermore, our tool is capable of filtering query records based on anomaly score (computed by our model), as shown in Fig. 3 which shows all queries with the anomaly score of 0.60 or more – anomaly score varies from 0 to 1, where 0 is least anomalous and 1 is the most anomalous. This value is calculated by the machine learning algorithm which is used while the classification of a query name. Additionally, our tool has a function to pause/resume the stream – if the user notices a malicious domain in real-time, then (s)he can pause the visualization stream to analyze the attributes of the malicious domain.

IV. CONCLUSION

In this demonstration, we have showcased our open dataset containing more than one million of malicious DNS queries (generated by our DNS exfiltration tool) with their corresponding attributes. We have also demonstrated our web-tool for visualizing our learning based detection of anomalous DNS queries (DNS exfiltration and tunneling from enterprise networks) which is operational in our campus network.

REFERENCES

[1] Efficient iP, "The Global DNS Threat Survey," Tech. Rep., 2017.

[2] B. Krebs. Deconstructing the 2014 Sally Beauty Breach. [Online]. Available: https://krebsonsecurity.com/2015/05/deconstructing-the-2014-sally-beauty-breach/

[3] J. Ahmed, H. Habibi Gharakheili, Q. Raza, C. Russell, and V. Sivaraman, "Real-Time Detection of DNS Exfiltration and Tunneling from Enterprise Networks," in *Proc. Integrated Network and Service Management (IM)*, Washington, DC, USA, April 2019.

[4] R. Qasim. (2018) DET (extensible) Data Exfiltration Toolkit. [Online]. Available: https://github.com/qasimraz/DET

[5] (2018) MasterCard Credit Card Generator. [Online]. Available: https://www.getcreditcardinfo.com/masterbulkgenerator.php

[6] (2018) Nozzle: DNS Exfiltation Data. [Online]. Available: https://nozzle-data.sdn.unsw.edu.au/dns-exfiltration-dataset/files/ExfiltrationAttackFQDNs.csv

[7] (2018) Nozzle: DNS Monitoring Tool. [Online]. Available: https://nozzle-data.sdn.unsw.edu.au/dnsQueries