CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Technical Disclosure Common

Technical Disclosure Commons

Defensive Publications Series

February 20, 2019

CROSS-DATASET MALICIOUS ACTORS IDENTIFICATION

Jan Jusko

Danila Khikhlukha

Follow this and additional works at: https://www.tdcommons.org/dpubs_series

Recommended Citation
Jusko, Jan and Khikhlukha, Danila, "CROSS-DATASET MALICIOUS ACTORS IDENTIFICATION", Technical Disclosure Commons, (February 20, 2019)
https://www.tdcommons.org/dpubs_series/1966

This work is licensed under a Creative Commons Attribution 4.0 License.
This Article is brought to you for free and open access by Technical Disclosure Commons. It has been accepted for inclusion in Defensive Publications Series by an authorized administrator of Technical Disclosure Commons.

CROSS-DATASET MALICIOUS ACTORS IDENTIFICATION

AUTHORS:
Jan Jusko
Danila Khikhlukha

ABSTRACT

Techniques are described herein for convicting malicious actors across datasets of different origins. The algorithm allows correlation of the available ground truth knowledge from one dataset with observations in another dataset. In the network/endpoint security field this algorithm allows for conviction of malicious network traffic and identification of Command and Control infrastructure of newly detected malware, even if no direct communication between binaries and domains is observed.

DETAILED DESCRIPTION

In the fields of networking and endpoint security there is a common task which can be defined as follows: “Given knowledge of some malicious binaries/domains and some relationships between them, identify additional malicious binaries/domains.” There are approaches that tackle only one side of this problem, either from the endpoint perspective or from the network perspective.

Both the endpoint and network telemetries may now be observed. Such cross-layer visibility opens up the possibility for additional inference tasks, such as: “Given the knowledge of some malicious domains, find malicious hashes related to the same threat (or vice-versa).” In practice, this task can be translated to solving very specific problems:

1. Given a set of malicious domains, identify malicious hashes that belong to the same malware campaign, i.e. attribution of observed malicious domains to specific malware campaigns.
2. Given a set of suspicious domains, confirm that they are malicious (if there are malicious binaries related to them).
3. Given a set of malicious binaries, identify malicious domains related to the binaries, thus uncovering the Command and Control infrastructure of the observed malware. This

Published by Technical Disclosure Commons, 2019

leads to creation of Indicators of Compromise (IOCs) that can be used in number of network-based security solutions.

4. Improving efficacy of endpoint and network security products by using known IOCs in either network or endpoint layer and projecting to the other layer.

The aforementioned task may be solved by a simple query, assuming one can observe direct connections from malicious binaries to malicious domains. Unfortunately, this is rarely the case, as most malware (especially with high severity) injects itself into legitimate processes. Then, no direct inference of maliciousness between binaries and domains may be performed reliably.

An algorithm is described herein that enables solving the cross-layer learning task effectively in a distributed fashion. In particular, the algorithm may convict malicious actors across network and endpoint telemetries. Given the access to the telemetry, the algorithm may identify a relationship between clients and domains observed in the network telemetry, as well as a relationship between clients and binaries observed in the endpoint telemetry. The algorithm assumes endpoint and network telemetries are collected over the same set of clients. The algorithm is described as follows for the specific case of known malicious domains, with the goal of identifying related malicious binaries.

Since the clients in both telemetry data sets are the same, both domains and binaries may be represented in the same space of clients. The algorithm may proceed as follows. First, a vector is created for each observed domain. The vectors have lengths equal to the number of clients observed in the telemetry data sets, and each element of such vectors corresponds to a client node. Elements representing clients connecting to the domain are equal to one, and the remaining elements have value of zero. Next, L2 normalization is applied for each vector.

Second, a vector is created for each observed binary. The vectors have lengths equal to the number of clients observed in the telemetry data sets, and each element of such vectors corresponds to a client node. Elements representing clients hosting the binary are equal to one, and the remaining elements have a value of zero. Next, L2 normalization is applied for each vector.

Third, a subspace is spanned by the set of vectors representing the ground truth (i.e., domains that are known to be malicious). A linear subspace may be created which contains all vectors associated with malicious domains. Fourth, the basis W may be found for the subspace using QR decomposition or a similar technique. Fifth, an orthogonal projection of vectors representing binaries to the hyperplane defined by W may be calculated. Sixth, the maliciousness score may be calculated as an absolute value of the projection vector.

Figure 1 below illustrates the intersection between datasets. Endpoint telemetry (B-C) and network telemetry (C-D) share the same set of clients. As a result of the vectorization procedure, some domain D and binary B are associated with vectors **D**={1, 1, 1, 0} and **B**={1, 0, 1, 1} respectively. After the L2 normalization, the vectors become **D**={1.732, 1.732, 1.732, 0} and **B**={1.732, 0, 1.732, 1.732}.

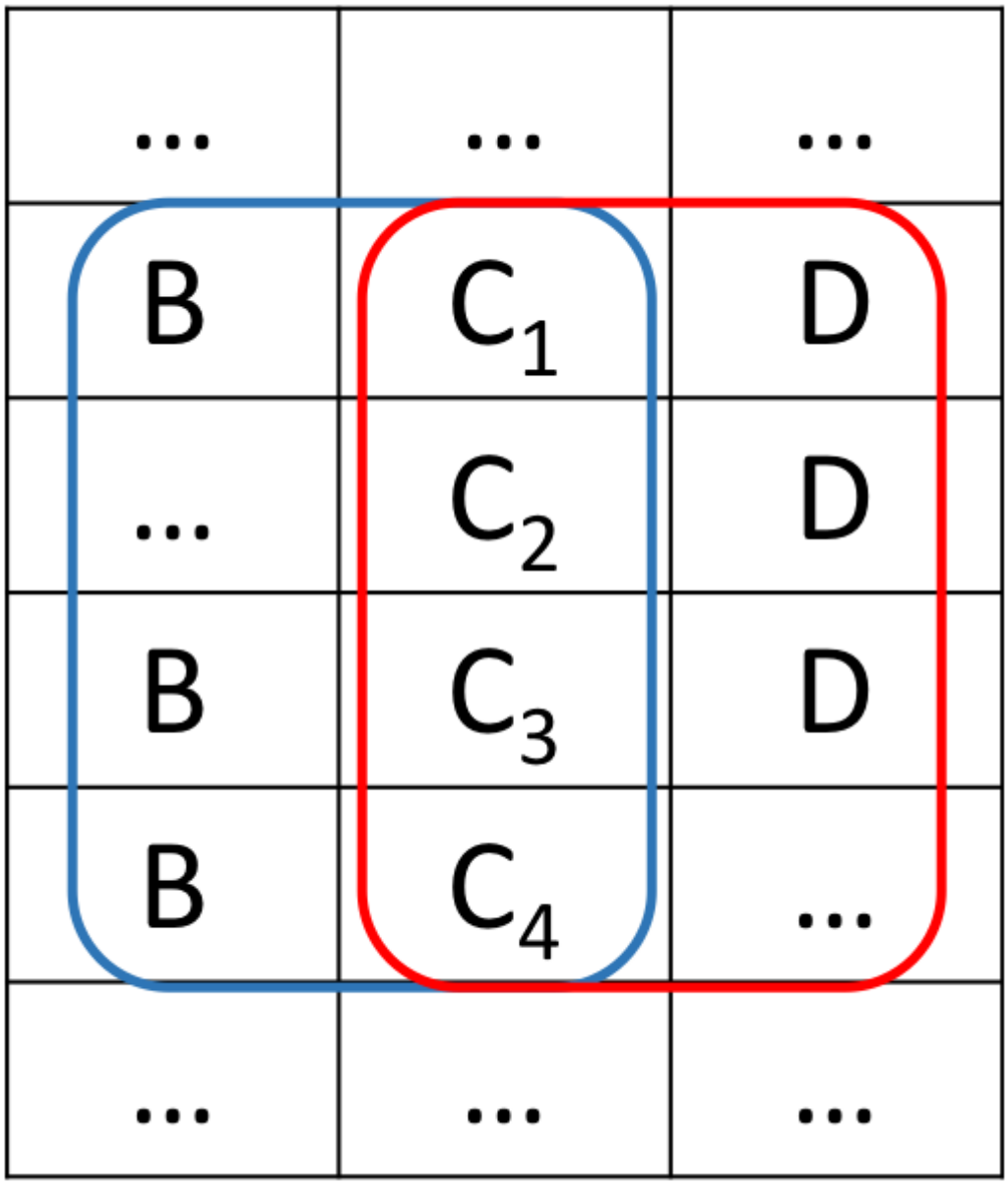

Figure 1

The obtained score may be treated as the probability of a binary being related to the infection that is causing the network traffic. Thus one can attribute the malicious network traffic to a specific malware family and/or convict suspicious traffic as malicious. The algorithm may be reversible: given the knowledge of malicious binaries, one might aim to identify malicious domains constituting the Command and Control infrastructure of a specific malware family.

This algorithm may be extended to any number of datasets. The only requirement for the datasets is to share some common entities that may be used as representation (e.g.,

use the same set of clients). Also, the algorithm may be used in a single domain. For example, given the knowledge of malicious domains, it may be possible to identify other malicious domains. The same goes for binaries. In fact, this algorithm may be used to efficiently solve a multitude of tasks.

In summary, techniques are described herein for convicting malicious actors across datasets of different origins. The algorithm allows correlation of the available ground truth knowledge from one dataset with observations in another dataset. In the network/endpoint security field this algorithm allows for conviction of malicious network traffic and identification of Command and Control infrastructure of newly detected malware, even if no direct communication between binaries and domains is observed.