REA605 ASSIGNMENT 5

Literature Review Draft 2

Gaston Carvallo - 048530133, Cedric De Pano - 110569159, Loyd Rafols - 022827158

REA605: Research Methodologies

February 15, 2021

Abstract

Recently, malware attacks have become more sophisticated, and the use of Domain Generation Algorithms (DGA) makes it harder to shutdown botnets and communications to and from malware, meanwhile the number of disruptive ransomware attacks is growing significantly (Cook, 2021). In this paper, we will look at previous research in the use of machine learning to detect malware attacks.

Contents

Introduction

Malware attacks are becoming more destructive, particularly the recent trend of ransomware attacks, where the ransom payment demands are increasing, and where attackers are not only threatens to lock data, but threaten to release it or sell it if payments are not made. As an example, Polish game developer CD Projekt Red was attacked by a group that infected their systems with ransomware. After CD Projekt Red refused to pay the ransom, the group created an auction for the source code it stole (Abrams, 2021), as well as not releasing the decryption key to undo the damage. In this paper, we examine the research on different methods to detect malware through the use of machine learning algorithms. First, methods that rely on network traffic for detection is reviewed. Since the number of malwares using encrypted traffic is growing (Cook, 2021), we will also look at some of the research that attempts to shed more light on how to detect these variants of malware. Finally, we reviewed articles that the attempts to detect domains generated by DGA one in plain text and one that uses encrypted traffic.

Background

Many organizations deploy an IDS[1] in order to protect themselves from malicious attacks. These devices seek to analyze data captured by its sensors and raise alarm to human analysts that make the final determination. Depending on the algorithm the IDS uses to determine if an attack is occurring, they can be categorized in on of 3 types (Gümüşbaş, et al, 2020):

[1] Intrusion Detection System - A software or hardware appliance that detects malicious traffic that passes through it.

a) Rule or signature based:

The IDS looks for specific data patterns that have previously linked to attacks. while fast and efficient they cannot detect novel attacks and are susceptible to adversaries attempts to obfuscate the attack by intentionally changing its pattern.

b) Statistic or anomaly based:

Statistics-based algorithms attempt to detect anomalies in user or network behavior by creating a model of normal behavior and using statistical analysis to determine if the observations are significantly different (Gümüşbaş, et al, 2020).

c) Machine learning (ML):

Consist in training a classifier, an algorithm that is trained to classify data in categories. most commonly attack/legitimate.

Das and Morris (2017) presented a list of machine learning (ML) methods and datasets used in cybersecurity, they highlighted the need to fully understand the data that will be used to train the selector - not only to be able to select the features to be incorporated in the model but to transform the raw data into a format that can be used by the ML algorithms. They also identified some of the datasets available to researchers, the DARPA 1998/99 datasets, the KDD cup 1999[2] and the Mississippi State University SCADA[3] dataset, which they used to analyze the effectiveness of some of the most common ML methods - specifically, Naïve Bayes, Random Forest, J48, and the OneR method. To test the effectiveness of the ML methods the receiver

[2] The KDD Cup is an annual data mining and knowledge discovery competition.
[3] Supervisory Control and Data Acquisition

operating characteristics curve was plotted. This is done by plotting the true positive rate vs the false positive rate (Das and Morris, 2017, 4).

Literature Review

Within the scope of this research topic, we split the state of the art along two main topics. The first relates to the state of detecting malware through analyzing the network traffic that is generated, both encrypted and unencrypted. The second topic is about the state of augmenting malware detection using machine learning.

Detecting malware through network traffic

Malware detection has traditionally been classified in static and dynamic analysis, where static analysis looks at the source code of the malware in isolation. Signature based detection is one of the main approaches to detect malware in this manner, while it can be fast and efficient for known malware it is ineffective for novel attacks and is susceptible to obfuscation attempts, like making changes to the source code or encrypting the file (Aslan and Samet, 2020, 6253) dynamic analysis in the other hand looks how the malware behaves, e.g., what system call it makes or how it changes the filesystem. While network analysis can be considered a subset of dynamic analysis, Manzano, Meneses and Leger (2018, 1) treat it as a separate category making a distinction between behavior that occurs within the host and the traffic analysis that occurs in the network. From within this subsection, we split the state of the art of the progresses done within the field of detecting malware through analyzing its unencrypted network traffic, as well as variants that use encrypted network traffic for communication.

Xia et al. (2020) proposed a Network-Assisted Approach (NAA) for detecting ransomware. In their proposed solution local detection and network-level detection are combined to provide users with a comprehensive report about each respective detection result. The local detection algorithm is applicable to many kinds of operating systems with a prototype for local detection targeting GNU/Linux systems as the current focus. The network-level detection adapted the any colony optimization algorithm and implements an ACO-based mechanism (ACOM). Additionally, they have implemented a Broadcasting Mechanism (BM) method for further data collection, using wisdom of the crowd, to help users determine their current safety state. The two proposed network-level mechanisms are separately suitable for detection in local and wide area networks. Characteristics of ransomware behaviour were analyzed, and common features showed obvious distinctions between safe and compromised hosts which will be used for their proposed detection method. Features such as keywords, function calls, data information, metadata information, and network traffic were used to judge if a host is in abnormal conditions. A brand-new feature referred to as “read/write pattern” is also considered for making accurate diagnosis on hosts. Most of the detection approaches pick several features and combine their checking results for false positives. The local detection algorithm they used, uses file entropy analysis, read/write frequency, and read/write pattern as the input parameters since it provides both high accuracy and efficiency. Wisdom of the crowd was used in the network-level detection to provide higher accuracy in detection results. Performance evaluation was done by building 100 Docker containers, which simulated network scenarios, where a hybrid ransomware sample such as GonnaCry infected hosts. Network-Assisted Approach was then launched in each infected container to evaluate for accuracy, message overheads, and latency.

Unencrypted network traffic from malware

Alhawi, Baldwin and Dehghantanha (2018) proposed a new model to detect ransomware on Windows machines called NetConverse. They analyzed 210 samples from 9 different ransomware families (Cerber, Cryptowall, Cryptolocker, CTB-Locker, Locky, Padcrypt, Paycrypt, Teslacrypt and Torrentlocker) and used 6 different machine learning classifiers (Bayes Network Multilayer Perceptron, J48, KNN[4], Random Forest and LMT[5]. Their experiment was made in three phases, first they collected samples for both ransomware and legitimate software from Virus Total. The final data set was created by extracting 9 features using TShark (a network protocol analyzer). Around 60% of the dataset was used for training and the rest to test the model efficiency. The machine learning mode was run on WEKA[6]. The J48 algorithm achieved the highest True Positive Rate (TPR) of 97.10% with the lowest False Positive Rate (FPR) of 1.60%.

Zhu and others introduce a network behavior-based method for detection of malicious reverse connections (Zhu et al, 2018). After concluding on a typical network communication pattern, network behavior features are extracted from TCP[7] sessions to be used as the detection model input. Algorithms were then applied on network traffic data collected, distinguishing malicious traffic from legitimate sessions, such as cloud applications and P2P.[8] The proposed method has proven to also work for encrypted malware traffic, specifically remote access trojans. Ability of handling imbalanced data sets were evaluated based on detection accuracies of the tested algorithms. Ghafir and others also presented an approach for detecting of botnet C2 traffic that is capable of real time detection (Ghafir et al, 2018). In their proposed system called BotDet,

[4] k-nearest neighbors
[5] logistic model tree
[6] Waikato Environment for Knowledge and Analysis
[7] Transport Control Protocol - used for transporting data over a network, the internet.
[8] Peer-to-Peer - direct connections from one host to another, without relying on a centralized host.

which is divided in two stages. They developed four modules for detection of different approaches used in malicious C2 communications. Additionally, to reduce the rate of false positives, they also designed a framework for correlation which balances false positives and true positive rates.

Encrypted network traffic from malware

In order to obfuscate their presence, some malware variants and families encrypt their traffic to make it harder to detect. In this subtopic, we will look at some of the work done to classify encrypted traffic. Jakob Premrn explores creating a device capable of detecting encrypted C2[9] channels using a machine learning model (Premrn, 2020, 5). Premrn's model presented a high False Positive Rate which would make his model unsuitable for day-to-day operations. He proposes that further work can be done to improve on this model by integrating it with cyber threat intelligence or some kind of whitelisting traffic (Premrn, 2020, 90).

Jaimin Modi explored detecting ransomware encrypted traffic by using certificate-based fields of the network connections captures (certificates, signers, and dates) as the features in the machine learning classifier model (Modi, 2019). The primary future work identified by Modi is to be more specific in its detection of malware, specifically the family of which a ransomware belongs to. (Modi, 2019, 68). Modi also proposes to increase his model efficiency in detecting ransomware C2 channels through the usage of generating random subdomains using a DGA.[10] In these two papers, there is a gap in regards to the different permutations of the features used to

[9] Also known as Command & Control - a method of controlling multiple infected hosts through a centralized server.
[10] Domain Generation Algorithm - a C2 channel used by malware to establish connections between an infected host and a controller server by way of a randomly generated subdomain name.

analyze malicious traffic - particularly with DNS requests and replies through DGAs, amongst other features - in the used machine learning model.

Modern malware tends to use DGAs to establish a channel to its C2 server instead of hard coded IPs this prevent defenders to block the specific IP or domain used by a family of malware. Zhang (2020) proposed a Deep Learning method to detect DGAs. This method is based on the assumption that domain names generated by DGAs are by their nature more random that benign domains and therefore should have a character distribution different enough that it can be detected (Zhang, 2020, 464). They test 5 machine learning algorithms, Support Vector Machine (SVM), Random Forest (RF), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN) and Long Short-Term Memory Networks (LSTM). They use Alexa top domains for the benign domain names, and use the UMUDGA[11] dataset that has over 30 million manually labeled DGA domains. For the regular machine learning algorithm (SVM and RF), 15 features were extracted from the dataset, for the deep learning algorithms the one-hot encoding method was applied (Zhang, 2020, 468). The deep learning models achieved over 0.95 precision in the binary categorization, that is, whether the domain was from a DGA or not.

However, because the DNS[12] protocol is not secure, there exist several protocols being studied to offer encrypted DNS services. Adversaries can then use these to avoid some of the DGA detection methods currently utilized. Patsakis, Casino and Katos (2019) developed IOCs[13] that could distinguish legitimate DNS from those generated by a malware DGA. They only looked at DNS over HTTP and TLS, on the assumption that other encrypted DNS protocols are

[11] University of Murcia Domain Generation Algorithm dataset located at https://data.mendeley.com/datasets/y8ph45msv8/1
[12] Domain Name System - a protocol used to resolve IP addresses into human-readable domain names.
[13] Indicator of Compromise

exotic enough that their presence alone can be detected and blocked by organizations firewalls or IDS (Patsakis, Casino and Katos, 2019, 4). They used Alexa top 1000 domains for their legitimate sample and created 1000 domain names from 10 known DGAs for their malicious traffic. Then captured the traffic using tcpdump[14] and extracted the features using tshark. They identified that the packet size of replies to DGA queries tend to be uniform while the Alexa domains presented higher variance (Patsakis, Casino and Katos, 2019, 6). Then studied the traffic as a time series and used the Hodrick-Prescott[15] filter to separate the trend, cyclical and error component, by identifying these 3 parameters they were able to attribute the traffic to their respective DGA (Patsakis, Casino and Katos, 2019, 7).

Malware Analysis using Machine Learning

In the current day and age, machine learning provides a boon for malware analysts, as it assists them in automatically detecting the changes made to a system by malware, as well as detecting permutations of a malware on different system configurations, or even different variants of a family of malware. Bae, Lee and Im explored using machine learning to detect and classify ransomware against benign operations or other types of malware, using Windows Native API[16] invocation sequences, more specifically those relating to file management and manipulation as their features used in the machine learning model, additionally by introducing a new indicator for classification (Bae, Lee and Im 2018, 4). Bae, Lee and Im used a dataset that contains the API invocation logs as their features, with 942 ransomware files, 830 malware files

[14] tcpdump is a Linux command line tool that can capture network traffic.
[15] The Hodric-Prescot filter is a statistical tool used in economics to separate cyclical components in time series.
[16] Application Programming Interface - a means for software to allow interaction with itself through predefined functions or tasks.

and 292 benign files (Bae, Lee and Im, 2018, 7). This model was use in conjunction with their indicator for classification model, called Class Frequency - Non-Class Frequency (CF-NCF), which focuses around how many times something shows up in a certain class, with a class being in this case, ransomware, malware or benign files (Bae, Lee and Im, 2018, 4-5). In conclusion, their model automatically generates a detection model using machine learning algorithms using their proposed indicator, CF-NCF, which they postulate can be iterated upon to detect new ransomware samples (Bae, Lee and Im, 2018, 10). We believe this paper is relevant because their scope of work only revolves around detecting ransomware through API function calls relating to file manipulation; that is to say, further work can be done to this detection model using different features such as network characteristics, or modifying the model to use with different variants or families of malware.

Noorbehbahani and Saberi explore malware analysis through detecting ransomware by using semi-supervised machine learning models. In their paper, Noorbehbahani and Saberi test the accuracy of their machine learning models with ten different ransomware families - Charger, Jisut, Koler, LockerPin, Pletor, PornDroid, RansomBO, Simplocker, Svpeng, WannaLocker - against benign datasets using five feature selection algorithms, Correlation-based Feature Subset Selection (CFS), OneR, Gain Ratio, ReliefF and Chi-squared (Noorbehbahani and Saberi, 2020, 2). The results found concludes that all five feature selection algorithms were moderately successful in their accuracy of determining malware versus benign files (generally in the range of 60-75%), with accuracy decreasing as additional validation folds[17] are conducted (Noorbehbahani and Saberi, 2020, 4-5). In conclusion, Noorbehbahani and Saberi claim that that utilizing Wrapper RF classification in combination with Chi-squared or OneR feature selection

[17] K-fold cross-validation is a means of validating data multiple times to mitigate biasing information.

models are the most effective at detecting ransomware in a semi-supervised environment. Noorbehbahani and Saberi identified a primary limitation within their research, such that the feature selection methods are supervised, so still requires human interaction to fully operate the machine learning model (Noorbehbahani and Saberi, 2020, 5). We believe that additional work can be done within detecting the same ransomware families in a supervised machine learning environment, or using a different dataset using the Wrapper RF classification with OneR/Chi-squared feature selection. There is possible work in performing research for ransomware detection in an unsupervised machine learning environment, but we believe, along with Noorbehbahani and Saberi, that this would be infeasible due to the poor accuracy around using unsupervised feature selection (Noorbehbahani and Saberi, 2020, 5).

Conclusions

In this paper, we reviewed the state of the art within detecting malware through analyzing the network traffic it generates, and detecting malware using supervised and semi-supervised machine learning models.

Within the realm of analyzing network traffic to detect malware, we reviewed several papers pertaining to two subtopics. The first subtopic revolves around detecting malware that uses unencrypted traffic as its means of communication. The second subtopic revolves around the same concept, but using encrypted traffic.

In our analysis of the literature for the subtopic of unencrypted traffic, work done by Zhu and others indicate that their proposed method works for detecting malicious traffic, and distinguishing that malicious traffic from legitimate traffic, such as cloud application and P2P

connections. Ghafir and others proposed a system, called BotDet, that is capable of detecting botnet C2 traffic in particular in real time (Ghafir et. al, 2018). Thus, the authors conclude that the state of analyzing unencrypted network traffic is well-defined, with most aspects of the art being covered by previous work. However, the authors believe additional work can be done by performing work in supervised machine learning models with different variants and families of malware. Additionally, different feature selection for datasets is also a consideration for future work that can be performed within the project.

Next, analysis of papers within the subtopic of encrypted traffic. Premrn explored creating a device that is capable of detecting encrypted C2 channels using a machine learning model (Premrn, 2020, 5). However, their model is infeasible in a day-to-day environment, due to the high False Positive Rate presented by the model. Modi and their work revolve around the concept of detecting ransomware encrypted traffic through using certificates. Patsakis, Casino and Katos developed a model for indicators of compromise, that can prove to help better distinguish legitimate DNS traffic from those generated from malware using a DGA. Modi proposes further work by improving their machine learning model to better specify the family or variant of ransomware identified, and to include randomly generated subdomains via DGAs as another feature to include in detecting ransomware C2 channels (Modi, 2019, 68). Additionally, Premrn proposes that further work can be done by integrating the model with cyber threat intelligence, or whitelists. With regards to this topic, we identified a knowledge gap around detecting malware using subdomains with encrypted communications, in particular those generated by a DGA.

Our other topic was reviewing papers that describe utilizing machine learning to analyze malware on a local, compromised host. Bae, Lee and Im used their proposed indicator for

classification to identify ransomware samples in a machine learning model. Noorbehbahani and Saberi explored detecting several different ransomware families through semi-supervised machine learning models. After reviewing these papers, we believe the state of the literature is missing a few aspects that could be better defined. In a supervised machine learning environment akin to that used by Bae, Lee and Im, different features can be selected such as network calls, function calls to generate a randomized domain name and other features not relating to files can be used as a basis of work for the project. Relating to work done by Noorbehbahani and Saberi, the work done only relates mainly to ransomware and its file modifications, rather than network traffic. Further work can be done by training the machine learning model around distinguishing malicious network traffic in ransomware to benign traffic from normal users in a semi-supervised environment.

Bibliography

Abrams, Lawrence, 2021. "CD Projekt's Stolen Source Code Allegedly Sold By Ransomware Gang". Bleepingcomputer, Last modified 2021. https://www.bleepingcomputer.com/news/security/cd-projekts-stolen-source-code-allegedly-sold-by-ransomware-gang/.

Alhawi, Omar MK, James Baldwin, and Ali Dehghantanha. 2018. "Leveraging machine learning techniques for windows ransomware network traffic detection." *Cyber Threat Intelligence*, pp. 93-106. Springer, Cham, doi.org/10.1007/978-3-319-73951-9_5

Aslan, Omar and Refik Samet, 2020. "A Comprehensive Review on Malware Detection Approaches," *IEEE Access*, vol. 8, pp. 6249-6271, doi: 10.1109/ACCESS.2019.2963724.

Bae, Seong Il, Gyu Bin Lee, and Eul Gyu Im. 2018. "Ransomware detection using machine learning algorithms." *Concurrency and Computation: Practice and Experience* 32, no. 18 (2020): e5422. doi: 10.1002/cpe.5422

Cook, Sam. 2021. "Malware Statistics In 2021: Frequency, Impact, Cost & More". Comparitech, Last modified 2021. https://www.comparitech.com/antivirus/malware-statistics-facts/.

Das, Rishabh and Thomas. H. Morris, 2017. "Machine Learning and Cyber Security," *2017 International Conference on Computer, Electrical & Communication Engineering (ICCECE)*, Kolkata, pp. 1-7, doi: 10.1109/ICCECE.2017.8526232.

Ghafir, I., V. Prenosil, M. Hammoudeh, T. Baker, S. Jabbar, S. Khalid, and S. Jaf,. 2018. “BotDet: A System for Real Time Botnet Command and Control Traffic Detection.” *IEEE Access*: vol. 6, pp. 38947-38958. doi: 10.1109/ACCESS.2018.2846740.2.

Gümüşbaş, Dilara, Tulay Yıldırım, Angelo Genovese, and Fabio Scotti. 2020 "A comprehensive survey of databases and deep learning methods for cybersecurity and intrusion detection systems." *IEEE Systems Journal*, doi: 10.1109/JSYST.2020.2992966.

Modi, Jaimin, 2019. "*Detecting Ransomware in Encrypted Network Traffic Using Machine Learning*", Master's thesis, University of Victoria.

Noorbehbahani, Fakhroddin, and Mohammad Saberi. "Ransomware Detection with Semi-Supervised Learning." In *2020 10th International Conference on Computer and Knowledge Engineering (ICCKE)*, pp. 024-029. IEEE, 2020.

Patsakis, Constantinos, Fran Casino, and Vasilios Katos, 2020. "Encrypted and covert DNS queries for botnets: Challenges and countermeasures." *Computers & Security 88.* doi.org/10.1016/j.cose.2019.101614.

Premrn, Jakob, 2020. "*Analysis of command and control connections using machine learning algorithms*." Master's thesis, University of Ljubljana, Faculty of Electrical Engineering.

Xia, T., Y. Sun, S. Zhu, Z. Rasheed, and K. Shafique. "Toward A Network-Assisted Approach for Effective Ransomware Detection." EAI Endorsed Transactions on Security and Safety: Online First, January 2021. doi: 10.4108/eai.28-1-2021.168506.2

Zhang, Yihang 2020."Automatic Algorithmically Generated Domain Detection with Deep Learning Methods," *2020 IEEE 3rd International Conference on Automation, Electronics and Electrical Engineering (AUTEEE)*, Shenyang, China, 2020, pp. 463-469, doi: 10.1109/AUTEEE50969.2020.9315559.

Zhu, H., Z. Wu, J. Tian, Z. Tian, H. Qiao, X. Li, and S. Chen. 2018. "A Network Behavior Analysis Method to Detect Reverse Remote Access Trojan." *2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS)*, Beijing, China, pp. 1007-1010, doi: 10.1109/ICSESS.2018.8663903.