International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-6, August, 2019

Leveraging Machine Learning Algorithms For Zero-Day Ransomware Attack

Sowmya Gaitond, Rekha S Patil

Abstract: ***Current global huge cyber protection attacks resulting from Infected Encryption ransomware structures over all international locations and businesses with millions of greenbacks lost in paying compulsion abundance. This type of malware encrypts consumer files, extracts consumer files, and charges higher ransoms to be paid for decryption of keys. An attacker could use different types of ransomware approach to steal a victim's files. Some of ransomware attacks like Scareware, Mobile ransomware, WannaCry, CryptoLocker, Zero-Day ransomware attack etc. A zero-day vulnerability is a software program security flaw this is regarded to the software seller however doesn't have patch in vicinity to restore a flaw. Despite the fact that machine learning algorithms are already used to find encryption Ransomware. This is based on the analysis of a large number of PE file data Samples (benign software and ransomware utility) makes use of supervised machine learning algorithms for ascertain Zero-day attacks. This work was done on a Microsoft Windows operating system (the most attacked os through encryption ransomware) and estimated it. We have used four Supervised learning Algorithms, Random Forest Classifier , K-Nearest Neighbor, Support Vector Machine and Logistic Regression. Tests using machine learning algorithms evaluate almost null false positives with a 99.5% accuracy with a random forest algorithm.***

Keywords: ***Ransomware, Malware analysis, Computer Security*, *Machine learning.***

I. INTRODUCTION

In present-day's digital connected world, agencies around the world are witnessing a rapid increase in cybercrime. As reliance on digital technology has increased, the economy has helped rebuild the business enterprise sector, but it has expanded to a variety of cyber attacks. Individual customers and businesses manage important documents, photos, reports, and organizational information in a digital format. These days, big- scale assaults have been completed using a type of malicious referred to as ransomware that turn down from access to person statistics documents and needs a huge extortion to be paid for redeem it. In a very short time, ransomware grew exponentially, becoming the most dangerous and highly competitive malware of all time. The attacks were done on numerous sectors such as finance ,coverage , banking , real property ,clinical, public administration to call some. Scareware is an early form of ransomware that exploits the false worries of a patient who fears that his machine will be infected with various viruses, adware and protection issues. Customers cheat to shop for fake antivirus products and pay the ransom to eliminate the infection. Person knowledge and superior protection software program have significantly reduced the chance of this kind of malicious viruses.Locker Ransomware uses the system's locked user interface to deny access to computer sources. He uses social engineering to threaten to pay the ransom to the consumer. As a result, tools and strategies are provided by many security companies, and you can recover the blocked personal interface to the maximum version. Encryption Ransomware Objective A file of information about a person with a specific extension that varies from parent to parent. File encryption using advanced encryption algorithms blocks access to personal statistics. In case of blame, the ransom is shown to the person who contains the threat message in order to completely remove the hostage file. The ransom is requested via bitcoin encryption. Device documents are not encrypted to keep the device running. Even after payment, you do not always receive the decryption key to restore encrypted documents. A large amount of extortion has created a brand new version of cryptographic ransomware ordinary. Most of the existing ransomware detection technology are known and powerful against very vulnerable samples for polymorphism, obfuscation and 0-day attacks which have been previously analyzed however extensively utilized in cryptographic ransomware. The indicators used for monitoring are much like, but similar to, regular malware, however do not completely seize the particular conduct supplied by using the ransomware product line. Several techniques and machine learning frameworks have been proposed and developed to detect transitional wear. However, the dynamic analysis technique has a limitation in that it can reconstruct a new variant of ransomware to reduce the detection rate by machine learning algorithms. Applying machine learning to dynamic machine inspection analysis can achieve detection rates of over 96%. Similarly, machine learning applications that analyze Android malware network traffic have achieved detection rates of over 99%. Because the Windows operating system accounts for approximately 89% of the desktop operating system market share, most of the current ransomware threats are targeted at personal computers running Windows operating systems.

Revised Manuscript Received on October 30, 2019.
* Correspondence Author
Sowmya Gaitond, Computer Science And Engineering, PDACE, Kalaburagi, India.
Rekha S Patil, Computer Science And Engineering, PDACE, Kalaburagi, India.

Here we focus on the Windows environment and analyze the .exe file for zero-day attacks and suggest using machine learning technology to detect Windows ransomware. The following are the contributions to achieving this goal.

(a) The evaluation study analyzed the .exe file and searched for a ransomware product family with captured data samples. These samples were collected using Virus Total.

(b) Evaluate four machine learning classifiers and analyze files as ransomware or benign.

(c) extracted and distributed the characteristics of the ransomware sample.

This paper is composed as follows: Section 2 deals with related tasks related to ransomware detection topics. Section 3 describes the research demonstration and the three steps involved. Section 4 presents experiences and discusses the results. Section 5 presents conclusions and discusses the future work.

II. RELATED WORK

Related obligations may be divided into broad categories. A) deal with ransomware as a network subset of all malicious programs and observe a familiar malware indicator for detection. B) a way in particular designed for ransomware based totally on its residences. Detection techniques based on antivirus signature.[1] are effective in opposition to recognized threats which are already present in the database however are prone to polymorphism and zero-day ransomware attacks. Institution coverage and appliances whitelist [2] are commonly used in small-business networks, however now not for people and the public zone.

There may be a dependency on the right maintenance of the whitelist software list. In addition, malware can still produce the most vulnerabilities in approved software. Static valuation detection techniques based primarily on control flow graphs [3], statistical drift graphs [4], and machine API calls [5] are suitable for obfuscation strategies, Polymorphic code And metamorphism. Malware and exploit tools have always been an important tool in the cybercriminal tool suite, and machine learning techniques have been used for decades to detect and analyze malwareMalware classification using machine learning is very effective in detecting malware on Android [7]. Ransomware steals the phone [8]. A formal method to detect the behavior of ransomware on the Android platform. We are getting inspiring results.Data set of 2,477 samples: 1 precision and 0.99 callback. It represents efficiency that surpasses the most popular commercial top ten ads.

The ransomware classifier based on the dynamic nature of the sample, EldeRan [9], has a TRUE (TPR) ratio of 96.3% and a low false positive rate (FPR) of 1.6%. Panorama [10] Capture information flow systems. During the experiment, 42 malware samples and 56 positive samples were evaluated. Panorama produces few false negatives and false positives. Unveil [11] is another device mastering based machine that detects ransomware the use of a pattern of lanceware that interacts with the os subsystem. Folks that done a true positive fee of ninety six. Three% (tpr) and a 0 false positive rate (rpf).

III. PROPOSED SYSTEM

This section introduces the experience of data collection, feature extraction, and machine learning classifiers. The three phases are described in FIG. 1

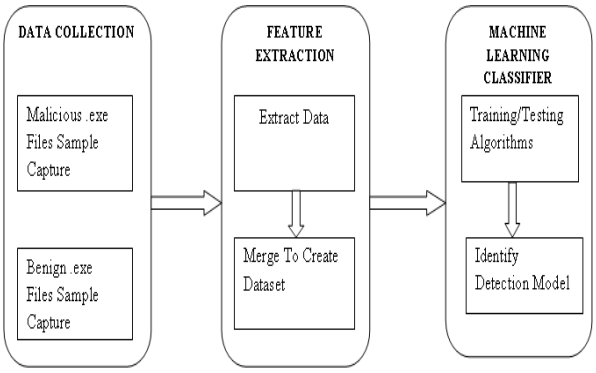

Fig. 1 Workflow showing the 3 phases for the demonstration

The first step is to collect data, Samples of .exe files are collected for malicious (ransomware) and good Windows operating system applications. The second step is to retrieve and merge related entities in the Entity Extraction step to create a dataset. The last step in the machine learning categorizer is learning and testing multiple algorithms in Python IDLE 3.6 (32-bit) machine learning to identify the optimal detection model.

III(A) Data Sample collection phase

This analysis focuses on Windows ransomware PE Files, these are samples of .exe file data that we compare with the characteristics of good quality applications that we use to and looks at the characteristics of all .exe files that are created when infected. An infected file / application spreads maliciously throughout the system and attempts to affect other applications / files. Therefore capture and use other classification techniques.

III(B) Selection and Extraction of Features

Function extraction was performed using the terminal function of the .exe file, which is a variety of application functions running on the system. Attributes Captures various antivirus .csv files that specify behavior and other attributes of the file. Each consign file is merged to create the initial preprocessed data.

III(c). Machine Learning Algorithms Classifier

In this experimental phase, we have identified a combination of the machine learning classifier and the function with the highest detection rate. The outline of classifier 4 is described in Table 1 we used.

Table 1

Machine Learning Classifiers

MACHINE LEARNING CLASSIFIER	PROS	CONS
Random Forest Algorithm	Can improve predictive Performance.	They are much harder and take more than decision trees.
Logistic Regression	It is very efficient, does not require too many computing resources.	We can't solve non-linear problems with logistic regression.
K-Nearest Neighbor	It is very simple to understand and equally easy to implement.	It is a lazy learner, i.e. it does not learn anything from the training data .
Support Vector Machine	It is useful for both Linearly Separable and Non-linearly Separable data.	Isn't suited to larger datasets as the training time with SVMs can be high.

IV(A) Evaluation metrics

IV. EXPERIMENTAL RESULTS

This part presents experimental results and evaluates the classifiers we've used to get the best search rates.
We have stated performance using six trendy measures: authentic positive charge (tpr), fake effective price (rpf), auc, precision, recall , and f1-score(F-Measure). The measurement effects are summarized in table 2.

Table 2 Evaluation metrics

METRICS	CALCULATION	VALUE
True positive rate (TPR)	TP/(TP+FN)	Correct classification of predicted malware
False positive rate (FPR)	FP/(FP+TN)	Goodware incorrectly predicted as malware
Precision	TP/(TP+FP)	Rate of relevant results
Recall	TPR	Sensitivity for the most relevant results
F-Measure	2 × (Recall × Precision)/(Recall + Precision)	Estimate of entire system performance

IV(B) Performance Results:

The time taken to construct each illustration in each segment of the experiments only one classifier (svm) , time it takes to construct every version in each step of the test. The alternative three classifiers spend much less time creating the models. Shown in table 3.

Table 3Comparison of processing time (in seconds)

Classifier	Processing Time
Random Forest Classifier	13.33
Logistic Regression Classifier	28.13
k- Nearest Neighbor Classifier	45.28
Support Vector Machine	1651.74

Table 4 summarizes the pleasant performance completed for each category. The results for each performance metrics dimension are displayed: tpr, rpf, accuracy, recall, f1 rating, auc, accuracy. The random wooded area classifier achieved the best performance out of the six measurements, and ninety nine. 5% is an important issue in growing the overall system performance fee (f-score).

The random forest classifier showed the highest accuracy with 99.7%, followed by logistic regression (65%), KNN (97.9%) and SVM (87.1%).

Table 4 Ransomware detection evaluation results

Classifier	Accuracy	TPR	FPR	Precision	Recall	F1-Score	AUC
Random Forest Classifier	0.997119	0.99981567	0.2276583	0.997049	0.996313	0.9953	0.997
Logistic Regression Classifier	0.650236	0.27760369	0.06391403	0.769152	0.277604	0.9721	0.978
K-Nearest Neighbors Classifier	0.979595	0.9959447	0.00282805	0.980841	0.971982	0.8579	0.886
Support Vector Machine Classifier	0.871089	0.97198157	0.01456448	0.771112	0.999816	0.6903	0.606

Figure 2 below shows a precise comparison graph for Random forest classifier, Nearest neighbor classifier K, Support vector machine, and Logistic regression, respectively.

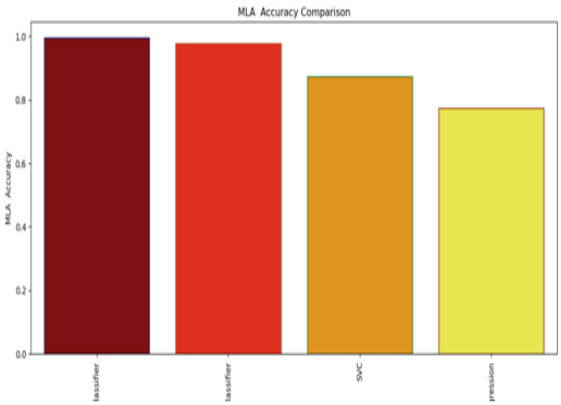

Fig 2 Machine learning algorithms accuracy comparison

V. CONCLUSION AND FUTURE WORK

Recent cyber security attacks from Cryptographic Ransomware have severely hindered businesses around the world. Based on an analysis of the complete Ransomware data set, this article provides an analysis of zero-day attacks that are exploited before the patch is released. Ransomware's behavior This demo is a quick fix for the Zero-Day Attack. Several computer-based classification programs are used to analyze Windows ransomware with multiple samples of PE file data. We used Different Supervised Machine Learning Classifier which is a affective against the evaluation process of zero –day ransomware attack. The selected classifiers were Random Forest, Logistic Regression, Nearest K-Neighbors and Support Vector Machine. Experimental results show 99.5% detection accuracy using the Random Forest classifier. This study provides a basis for further research by which researchers can elongate and develop data sets to include other ransomware products and extract additional functionality to enhance the detection process.

REFERENCES

1. SentinelOne,"The Truth About Whitelisting , "Dec 2014. [Online]. Available :https://sentinelone.com/2014/12/07/ the-truth-about-whitelisting.
2. P.Faruki, V.Laxmi, M.S.Gaur , and P.Vinod , "Mining control flow graph as api call-grams to detect portable executable malware," in 5th International Conference on Security of Information and Networks. ACM,2012,pp.130–137.
3. T.Wüchner, M.Ochoa , and A.Pretschner , "Robust and effective malware detection through quantitative data flow graph metrics, "in International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment. Springer,2015.
4. Y.Ye, D.Wang, T.Li,D. Ye, and Q.Jiang ,"An intelligent pe-malware detection system based on association mining," Journal in computer virology,2008.
5. M. Hopkins and A. Dehghantanha, "Exploit Kits: The production line of the Cybercrime economy?," in 2nd International Conference on Information Security and Cyber Forensics, InfoSec 2015, 2016, pp. 23–27.
6. A. Feizollah, N. B. Anuar, R. Salleh, and A. W. A. Wahab, "A review on feature selection in mobile malware detection," Digit. Investig., vol. 13, no. March, pp. 22–37, 2015.
7. N. Milosevic, A. Dehghantanha, and K. K. R. Choo, "Machine learning aided Android malware classification," Computers and Electrical Engineering, 2016.
8. F.Mercaldo,V.Nardone,A.Santone,andC.A.Visaggio,"Ransomware steals your phone . formal methods rescue it ," in International Conference on Formal Techniques for Distributed Objects , Components ,and Systems. Springer,2016,pp.212–221.
9. D. Sgandurra, L. Muñoz-González, R. Mohsen, and E. C. Lupu, "Automated Dynamic Analysis of Ransomware: Benefits, Limitations and use for Detection," no. September, 2016.
10. H.Yin, D.Song, M.Egele ,C.Kruegel , and E.Kirda ,"Panorama: capturing system-wide information flow for malware detection and analysis," in Proceedings of the 14th ACM conference on Computer and communications security. ACM,2007,pp.116–127.
11. A. Kharaz, S. Arshad, C. Mulliner, W. Robertson, and E. Kirda, "UNVEIL: A Large-Scale, Automated Approach to Detecting Ransomware," Usenix Secur., pp. 757–772, 2016.

AUTHORS PROFILE

Sowmya Gaitond Mtech student in Computer Network and Engineering at Poojya Doddappa Appa College of Engineering, Kalaburagi, Karnataka. Her Research interest is in the area of Computer Networks and Machine Learning.

Rekha S Patil M.Tech, Assistant Professor, Computer Science and Engineering Department at Poojya Doddappa Appa College of Engineering, Kalaburagi, Karnataka, She have teaching experience of 14 years. Areas of research are computer network and network security. She has published number of research papers in International and National journals and conferences.