Insights of Defending Encryption Attacks(Ransomware)

1st Yogesh Prabhakar
Department of Computer Science
Central Michigan University
Mount Pleasant, Michigan
email: prabh1y@cmich.edu

2nd Anusha Arjun Ganesh
Department of Computer Science
Central Michigan University
Mount Pleasant, Michigan
email: arjun1a@cmich.edu

3rd Kalyan Chakravarthy Narne
Department of Computer Science
Central Michigan University
Mount Pleasant, Michigan
email: narne1k@cmich.edu

Abstract—**Encryption attacks are gaining more attention in recent days, Ransomware is one of them. These attacks end target is to encrypt the victims/clients data and stop them from accessing. The attackers launching encryption attacks use different techniques to hijack/steal the victims/clients files and resources to demand ransom in exchange to free the encrypted/captured data or resources. Though there are variety of malware that are challenging today's business infrastructures, Encryption attacks like ransomware are true nightmares to everyone just because it impacts data and Data is ”Precious Wealth” for both attacker and defender. There are so many levels involved in the these kind of attacks. While looking into every step on how these attacks are performed we will simultaneously look more into how we secure ourselves from these kind of attacks. Every aspect of defending model is developed based on various attacking factors. The Defending model that is proposed here is concrete enough, not only for Ransomware attacks but for most of the attacks that are based on data encryption.**

I. Introduction

Encryption attacks are malicious software that stealthily encrypts user files and demands a ransom to provide access to these files. Several prior studies have developed systems to detect ransomware attacks by monitoring the activities that typically occur during a ransomware attack. Unfortunately, by the time the ransomware is detected, some files already undergo encryption and the user is still required to pay a ransom to access those files. Furthermore, ransomware variants can obtain kernel privilege, which allows them to terminate software-based defence systems, such as anti-virus. While periodic backups have been explored as a means to mitigate ransomware, such backups incur storage overheads and are still vulnerable as ransomware can obtain kernel privilege to stop or destroy backups. Ideally, we would like to defend against ransomware without relying on software-based solutions and without incurring the storage overheads of backups.

There are many attacks and defences, that emerge once we have large numbers of machines networked together. These depend on a number of factors, the most important of which are the protocols the network uses. A second set of factors relate to the topology of the network. Since the operations performed by ransomware are indistinguishable from benign software, ransomware can easily bypass various antivirus, making it increasingly prevalent in cyber criminals [2]. According to a study from IBM Security [1], the number of users who came across encryption ransomware in 2016 increased by more than 6,000 per cent over the previous year. The ransomware attacks cost their victims about a billion dollars in 2016 which is a 41x increase compared to the cost in all of 2015 [3]. To counteract ransomware, researchers have proposed several detection systems that use file access patterns [4, 5] or features of cryptographic algorithms [6] to identify ransomware. However, these detection mechanisms still cannot prevent ransomware from locking up user data. First, existing ransomware detection occurs only after observing the actual damage. Given that the encrypted data may contain the files considered to be valuable, victims still have to shoulder the burden of paying the ransom. Second, some ransomware can run with administrator privileges, which allow them to load kernel code and carry out kernel-level attacks. Given that the existing defense systems typically run within the kernel, ransomware can easily disable or work around the aforementioned detection mechanisms. To address these issues, one instinctive solution would be to enable file backup on local persistent storage (e.g., journaling and logstructured file systems [7, 8]) or remote machines (e.g., NFS [9] and cloud-based storage [10]). However, this is insufficient at guarding against ransomware. First, any file backup mechanisms inevitably impose storage overhead. Second, ransomware may find and jump to the backup and encrypt it regardless of whether it is on shared network drives, local hard disk drives, external storage devices, or plugged-in USB sticks. Third, ransomware with the kernel privilege can also terminate backup processes, making them futile against ransomware defense.

Machine Learning (ML) is gaining popularity in malware detection as it has proven to identify not only existing malwares but also new and obfuscated malwares (Rieck et al., 2011). There are two phases of work in malware analysis using ML (Handa et al., 2019). In the first phase called ‘Training Phase’, the ML algorithms formalize the set of features extracted from malicious and non-malicious data to build a predictive model. The second phase of ML algorithms is the testing phase, where the predictive model derived from the training phase is used to predict the benign behaviour of the malware.

II. Encryption Attacks

There are various kinds of malware revolving around us in real time. Few of them targets your data and tries to either steal or encrypt your data. Security of file contents is one of the most important problems for businesses in the ever-changing world of data and information transfer. Some information (emails, logins) can be password protected, but other information transferred by email or FTP is inefficient if protected by a keyword. This is where file encryption comes into play, providing the security and convenience that parties involved in file transfers are looking for.[11]

The process of encoding data is known as encryption in cryptography. This procedure turns plaintext data into ciphertext data. Only authorized parties should be able to decrypt ciphertext and access the original data. These kinds of attacks are generally called as Encryption Attacks.

Cyber-adversaries have devised new sorts of attacks in reaction to data encryption at rest. Cryptographic attacks,[12] stolen ciphertext attacks,[13] attacks on encryption keys,[14] insider attacks, data corruption or integrity attacks,[15] data destruction attacks, and ransomware attacks are some of the more recent dangers to data encryption at rest.

Some of these threats are countered by data fragmentation[16] and active defense[17] data security systems, which distribute, move, or mutate ciphertext to make it more difficult to detect, steal, corrupt, or destroy. [18]

III. What is Ransomware ?

As we discussed about encryption attacks, Ransomware is just another attack or it is a type of malware that demands money in exchange for access to a stolen function, usually data. Malware of this type has been identified as a major danger to computer and network security all around the world [19].

Cryptovirology's cryptoviral extortion assault, which holds the victim's data hostage, or cryptovirology's leakware attack, which threatens to reveal the victim's data, is installed discreetly on a victim's device. Critical data, which is vital to both individuals and businesses, is the true objective of this type of attack. In fact, because of the delicate nature of the malicious applications, the attack has expanded to mobile devices, making mobile malware detection technologies ineffective. [20] The quantity of online apps and services, as well as smart mobile devices, has grown tremendously in the last six years, resulting in a major surge in digital extortion [21]. Ransomware's influence has grown to the point where it is now considered the largest cyber fraud to affect organizations [22]. Around 80 percent of ransomware assaults target Flash vulnerabilities that companies should have fixed. Destructive ransomware can spread on its own, enslaving entire networks (i.e. businesses). This much-hyped genre, dubbed ransomware by the media, can't seem to stay out of the headlines. Despite the fact that it has not been extensively spread, the attack is worthy of note due to its audacity. The ransomware malware infects the user's machine using a variety of methods, including sending the victim a convincing email persuading them to open the file. When the trojan is placed on the victim's machine, it starts obfuscating or encrypting potentially important data such Microsoft office files, photos, archives, password files, and anything else the victim is likely to value highly and be willing to pay to recover. A ransom note is left in a prominent area for the user, such as a text file in the same folder as the now-inaccessible "ransomed" material. According to the ransom message, the victim can quickly regain access to the locked-up digital items if he or she contacts the author for money or makes a payment to a certain account.[23].

In these two years (2016 and 2017) the number of infections increased too. The most important ransomware in terms of infections in 2016 was Tesla Crypt [24], which reached 90 percent of all ransomware infections. In May 2016, its developers published the decryption key and the ransomware disappeared. In terms of profits, Locky was the most profitable ransomware in 2016, with more than 7 million dollars[25]

In 2017, there were significant infections in businesses, such as the WannaCry attack (May 2017) [26], which affected a number of major companies, including Telefónica, a Spanish telecommunications corporation. This ransomware version was still active (with minor upgrades) in 2018, affecting some of the UK's most major hospitals and costing the British National Health Service around £92 million [27]. Cerberus lawsuit was also highly important in that year, garnering 6.9 million dollars[25]. There were more than five different versions of this ransomware, according to the ransom note left on the infected system by the hackers. Other ransomware variants, including Ryuk, BitPaymer, GandCrab, and BitPaymer, targeted some of the world's most important businesses in 2018 and 2019 [28]–[30].

IV. Simulating Ransomware attack

In order to simulate ransomware, we need an isolated sandbox or a virtual machine. We are using Python script in order to simulate Ransomware attack on Ubuntu operating system. This attack is carried in three phases, Creating Keys, Encrypting and Decrypting files. Hence this simulation is just to identify the process that are initiated during encryption of the attack and build a defending mechanism which would be applicable on any encryption attacks as this defending model is based on identifying and killing processes that are responsible for encryption.

Creating a Key

The attacker would run a script on his machine that creates two asymmetric keys RSA Key's[31]. A private key and public key will generate once the attacker runs the script. The attacker then creates a ransomware script that will be available on the target machine that encrypts the all the files, encrypts the system and also decrypts the system. To do this we used the following encryption module.

```
from cryptography.fernet import Fernet
```

Encryption

The public key then used to encrypt the symmetric fernet key[32], that encrypts and decrypts the files on the target machine. The attacker then packages the ransomware with the public key and send it off to which ever target machine or victim that he is trying to get some ransom off so that the attacker will pretend that he just packaged it up. The attacker may send the ransomware package via mail or some other online platforms. If the target victim download the ransomware package that was sent to him by the attacker, his machine will get encrypted. Even if the victim knows the public key he can't decrypt his machine, as the public key can go out to many machine it doesn't make any difference as long as the attacker keeps the private key private. The attacker keeps the private key for himself.

Decryption

The Ransomware encrypts the victim's system. The ransomware encryps thoses file with the specified extension mentioned in the ransomware script. If in the ransomware script only .txt are included then only .txt files on the target machine will encrypt. Now the victim has some ransomware on his computer. If the files on the victim device got encrypted, a RansomNote.txt file will be available on the victim's device. In the RansomNote.txt the attacker will mention about the steps to be followed to get the decrypt files. Also fernetkey.txt will generate which contains the encrypted fernetkey.

V. Our Defending Model

In our Defending model, we are securing our system from malware by monitoring the system processes dynamically and if any unusual process is found which is not expected to run on the system, then it is killed.

We've build our Defending model in such a way that if our defending model program is running then it would take care of any process which is running with irregular commands that uses system resources in a peculiar format i.e., sudden growth in CPU usage or Memory usage or both and would terminate the process. If there is any expected usage of resources, then in a planned window of time our model can be stopped.

Therefore to develop this Defending model, we are going with Machine Learning approaches to make predictive analysis in the system.

Machine Learning

Machine learning (ML) is a type of artificial intelligence (AI) that explored the possibility to increase prediction accuracy without being specifically designed to do so. In order to forecast new output values, machine learning algorithms use past data as input [33].

Machine Learning (ML) is getting popular in malware detection due to its ability to detect not only current malware but also novel and veiled malware.

Generally there are three phases of work in Machine Learning.

1.Collecting the data, Pre-process the data according to the requirement i.e., to be able to label the data properly. This phase is generally called as data-set preparation phase.

2. The ML methods formalize the collection of features acquired from harmful and non-malicious data to develop a predictive model in the second step, referred to as the 'Training and testing Phase'. The testing step of ML algorithms is where the predictive model created during the training phase is utilized to forecast the malware's benign behavior with test data-set.

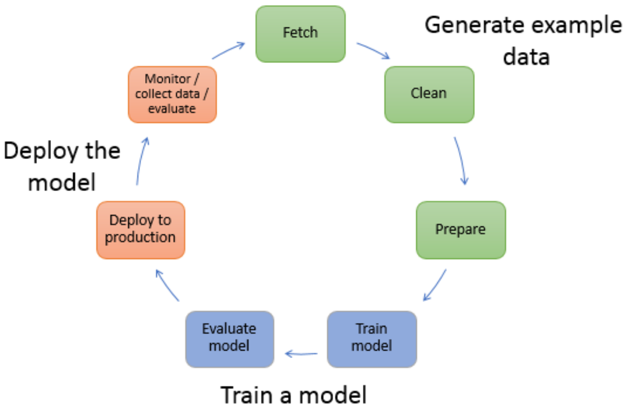

Fig. 1. General Workflow of Machine Learning.

3. Once the test is successful, here comes the deployment phase where the developed model is implemented in the system and real-time data is given as input dynamically and the developed code should be able to detect the labelled data if found.

Why Machine Learning ?

Before we get into why we are using Machine Learning, we need to understand our goal here, the goal in developing our defending model is to automatically identify the malware or a faulty process which is not expected to run on your system and once it is identified it has to be terminated.

In this scenario, if we use a naive way to kill a process just by its name/command would over-fit only to the malware that is used in the simulation but that is not our ultimate goal. Not only by command name, if we just use disk activity monitoring or CPU monitoring or Memory or whatever parameter we consider it would never serve the purpose of our defending model.

So We've chosen Machine Learning for our Defending model to overcome All these factors and to server our purpose we need to consider all the parameters like unusual commands, usage of CPU, usage of memory etc at once and need to understand the usual behaviour of your system in order to differentiate unusual behaviour. This can be easily achieved using Machine Learning.

In Machine Learning we have variety of approaches for predictive analysis, where they can be categorized as supervised and unsupervised machine learning.

Supervised Machine Learning

Machine learning is the process of learning a set of rules from a training set of examples, or more broadly, developing a classifier that can generalize from new instances. The learning techniques are as follows: the first step is to gather the data set; if a data set obtained by any of the arbitrary methods is not immediately suitable for induction, then next step is to collect the data set. This may have noisy or lacking data values, necessitating extensive prepossessing [34].

To train the machine, you require well-labeled data in supervised learning. It means that some data has already been labeled with the correct response. It is comparable to learning that occurs in the presence of a supervisor [35].

The feature subset selection stage is the act of detecting and deleting as many unnecessary and redundant features as possible [36]. As a result, the dimension of data is reduced, allowing algorithms to operate more quickly and efficiently. However, many characteristics are interdependent and may affect the accuracy of supervised Machine Learning classification models.

Unsupervised Machine Learning

Unsupervised learning is a type of machine learning in which the model is not supervised. Alternatively, you should let the model figure out what it needs to know on its own. It is mostly concerned with large data-sets [37].

Why Supervised Machine Learning ?

In our model, the output after malware detection is to terminate the processes associated with the malware. To achieve this we need to go with supervised Machine learning as the model training and behavioural study of system is easily achieved.

Supervised learning enables for the collection of data and the output of data from prior experiences. In supervised learning, models are taught to produce the desired output using a training set.

This training data-set includes both correct and incorrect results, which allows the model to improve over time. The loss function is used to evaluate the algorithm's correctness, and it is changed until the error is minimized to the desired level.[37]

Under supervised Machine Learning, classification algorithms is the way to reach our goal as we need to classify the good processes and malware processes. Hence we will be using classification algorithms.

Classification Algorithms

Classification uses an algorithm to accurately assign test data into specific categories. It identifies certain entities in the data-set and makes educated guesses about how those entities should be labeled or defined. k-nearest neighbor, Linear classifiers, decision trees, support vector machines (SVM), and random forest are some of the most common classification algorithms.[37]

1. Decision Trees

By learning simple decision rules inferred from past data, the purpose of employing a Decision Tree is to develop a training model that can be used to predict the class or value of the target variable (training data). We start from the root of the tree when using Decision Trees to predict a record's class label. The root attribute's values are compared to the record's attribute's values. Based on the comparison, we jump to the next node by following the branch that corresponds to that value.[38]

2. Random Forest technique:

Random Forest is a well-known machine learning algorithm that uses the supervised learning method. It can be used for both classification and regression problems in machine learning. It's based on ensemble learning, which is a technique for combining multiple classifiers to tackle a complex problem and improve the model's performance. ”A Random Forest is a classifier that combines a number of decision trees on different subsets of a dataset and averages them to increase the dataset's predicted accuracy.” Instead, relying on a single decision tree, the random forest collects the forecasts from each tree and predicts the final output based on the majority votes of predictions.[39]

3. Support Vector Machines (SVM):

One of the most advanced algorithms for classification jobs is the Support Vector Machine (SVM) [40]. Support Vector Machine (SVM) is a relatively simple Supervised Machine Learning Algorithm used for classification and/or regression. Its learning phase entails solving a Convex Constrained Quadratic Programming (CCQP) problem to find a set of parameters, for which numerous successful algorithms have been proposed [41].

However, finding the best parameters does not finish the SVM's learning phase: a collection of extra variables (hyperparameters) must be modified to find the SVM with the best performance in classifying a certain set of data. This phase is commonly referred to as model selection, and it is intimately tied to the estimation of a classifier's generalization capacity (i.e., the error rate achievable on new and previously unseen data), as the chosen model is defined by the minimum estimated generalization error.

Unfortunately, fine-tuning the hyper parameters is a difficult task that remains an open research question. [42-45]. It is more preferred for classification but is sometimes very useful for regression as well. Basically, SVM finds a hyper-plane that creates a boundary between the types of data. In 2-dimensional space, this hyper-plane is nothing but a line.

In SVM, each data item in the data-set is plotted in an N-dimensional space, where N is the number of features/attributes in the data. Next, determine the best hyperplane for separating the data.[46]

Algorithm Selection

The selection of a learning algorithm, in particular, is a crucial task. When the preliminary testing is completed and the results are favorable, the classifier is generalized [47]. There are a variety of methods for calculating the accuracy of a classifier.

One method is to divide the model set into one- forth training set and rest for test set. Another approach is cross-validation, which divides the training data set into mutually exclusive equal-sized subsets and trains the classifier on the union of the remaining subsets for each subset.

Therefore considering the availability of data and other factors we've chosen to go with SVM Algorithm.

Why SVM ?

For two-group classification issues, a support vector machine (SVM) is a supervised machine learning model that uses classification techniques. SVM models can categorize new text after being given sets of labeled training data for each category.

They have two key advantages over newer algorithms like neural networks: greater speed and better performance with a limited number of samples (in the thousands). This makes the approach ideal for text classification problems, where it's typical to have many classifications.[48].

In machine learning, there are a variety of algorithms for classification, but SVM is superior to the majority of them due to its higher accuracy and it is more effective in high dimensional spaces. SVM has high ability to predict unseen data, compared to other machine learning Algorithms.[49]

Lets understand in detail how exactly does SVM work before using it in our model.

How does SVM work ?

The working premise of support vector machines is straightforward: create a hyperplane that divides the data-set into classes as shown in Fig2. Let's start with a hypothetical scenario. Assume you have to distinguish red triangles from blue circles in a given data-set.

Your goal is to draw a line that divides the data into two categories, separating the red triangles from the blue circles.

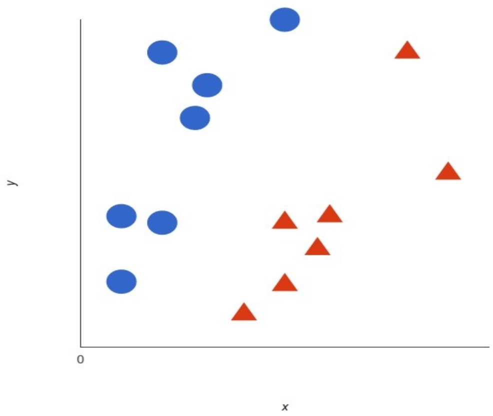

Fig. 2. SVM Plane with classified data

While there may be an obvious line that separates the two classes, there may be several lines that potentially do this. As a result, you cannot agree on a single line that can do this duty. We must locate the points that are closest to both classes, according to SVM. Support vectors are the names given to these points.

The next step is to determine how close our dividing plane is to the support vectors. The margin is the distance between the points and the dividing line. The goal of an SVM algorithm is to increase this margin as much as possible. The hyperplane becomes the best when the margin reaches its maximum.

By establishing a well-defined decision border, the SVM model attempts to widen the gap between the two groups. Our hyperplane partitioned the data in the example above. Our data was two-dimensional, but the hyperplane was just one-dimensional.

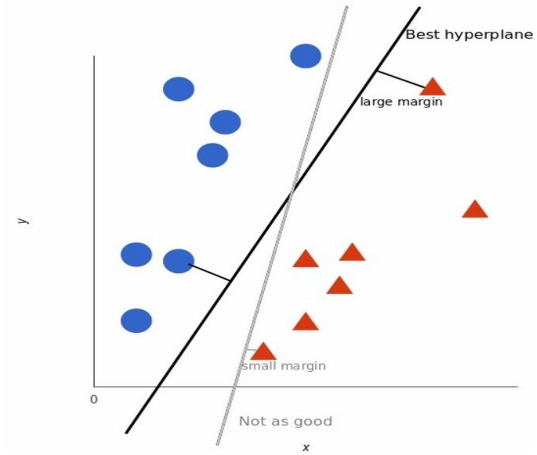

Fig. 3. SVM Hyperplane

We have an n-1 dimensional subset for greater dimensions, such as an n-dimensional Euclidean Space, that separates the space into two disconnected components[50-52].

VI. Mechanism of our Defending Model

In our Defending model initially we need to choose a data-set. Then Train your model with labelled data-set and then finally develop and implement a rigid model to detect malware.

Data Set

To identify the malware in a system, firstly we need to understand the system and its usual behaviour. So we need to understand its genuine processes that run all the time. The most effective way to observe or monitor a system is its top command.

The top command in UNIX based systems will have all the currently running processes information. So we are considering the top output as our data set.

In our data set we are considering five features %CPU, %MEM, COMMAND, PR(Priority of the process) and NI(Nice Value). So here the the idea behind selecting these parameters as classification features is to identify the process with irregular process name and its CPU usage and memory usage and also its priority and nice value, so that we can understand the behaviour of malware.

Then we need to proceed with data cleaning, pre-processing and labelling the data so that we can train our model and this will be done in Training phase. We will look in detail regarding how do we develop our training model.

Training Data Set

Firstly we need data set(top command output) as input to our train model code. Once the data is ready we can start developing the training model. We need to import below libraries.

```
  import pandas as pd
  from sklearn.preprocessing import
OneHotEncoder
  from sklearn.preprocessing import
LabelEncoder
```

```
from numpy import array
from numpy import argmax
```

Pandas is an open-source data analysis and manipulation tool built on top of the Python programming language that is fast, powerful, flexible, and simple to use.[53] Pandas is based on two core Python libraries: matplotlib for data visualization and NumPy for mathematical operations. Pandas acts as a wrapper around these libraries, allowing you to use fewer lines of code to access many matplotlib and NumPy methods. Pandas'.plot(), for example, combines multiple matplotlib methods into a single method, allowing you to plot a chart in a few lines. Prior to pandas, most analysts used Python for data munging and preparation before switching to a more domain-specific language like R for the rest of their workflow. Pandas launched two new types of objects for storing data that make analytical tasks easier and eliminate a need to change tools: Series, which have a list-like structure, and DataFrames, which have a tabular structure.[54]

The sklearn.preprocessing package includes a number of common utility functions and transformer classes for converting raw feature vectors into a representation suitable for downstream estimators.[55]

The next thing to consider is data cleaning, which includes null values, NA, or None. For cleaning the data, we are using the dropna()[53] method. Here we are not dropping the duplicates because, in real time, whatever process we get, it should give us an output. But we dropped the duplicates in the training model file. For machine learning, we have to send the input in vector form[56], but our input is not in vector form, it is in text form for the command column.

So first we will go through Label Encoding, this approach is very simple and it involves converting each text value in a column to a number. Once we have a number in the column instead of text then we can use One-Hot Encoding, One-hot encoding is the process of converting categorical data to numerical data for use in machine learning. Categorical features are converted into binary features that are "one-hot" encoded, which means that if a feature is represented by that column, it receives a 1. Otherwise, it receives a 0.[57]

To train the data, we have to label the data with classification of normal processes and malware processes. We will be Storing the features in the X variable and the label in the Y variable. We shall split the data into the training and test sets. With test_size = 0.25, 25% of the data will be kept as the test set, while the remaining 75% will be used for training as the training set. [58][59]

Now we have to build the support vector machine classifier. For that, import the SVC classifier from sklearn.svm. By using the fit () method, we are training our model. We are sending both train_features and train_labels to the fit method. Here, train_features are scenarios, and train_labels are expected output. Using the predict method, we can calculate the accuracy.

Finally after training the model, we will save the model into .sav file as we won't train the model again and again. from next time we will load the model from the disk instead of training the model every time by calling .sav file in the final code.[60]

Developing Our Defending model

Initially we will be importing all the required libraries as mentioned in Training data set and also below libraries additionally.

```
import numpy as np
import os, signal
import joblib
```

Then reading the input (i.e., topfile.txt in our case) and creating the data-frame using pandas.

Then we will clean and prepare the data as we did while training data-set.

Now, Once the data is ready for evaluation, we will be loading the trained model from disk(i.e., .sav file which we saved after training the model) using joblib library.

With the trained model, our defending model will start evaluating the data and start classifying the data. Here we need to Develop our code in such a way that if the data that was labelled as malware is identified then it need to further fetch the process id of the process and terminate using OS, signal libraries.

VII. Implementation

To implement our final code in the system, we need to parse the top output dynamically to our final code. So we need to make sure two things, we need to parse the top output continuously to our final code and our code need to be up and running as long as the top command is producing data.

Hence we've chosen bash scripting to fulfill both the requirements. In bash scripting we are using a infinite while loop for appending the top command output to a text file in batches and also in the same loop we are running our developed Defending model to identify if there is any malware in those processes.

Moreover, as the whole implementation is in infinite loop, it is up to user to run or stop our defending model. Also the bash script can be run in the background so that our defending model can be running all the time.

VIII. Conclusion

Encryption attacks can be deadly and dangerous as the dependency on data is huge in recent times and with increasing dependency the risk of cyber attacks are also in the race. Our Defending model is just an unique approach to help your system to fight against malware. Though there are few areas where enhancements or improvements can be done to be more diverse while detecting malware but still our defending model has fulfilled its purpose. It will be efficient enough if it is retrained to identify any kind of malware. The more you train the more efficient results can be expected.

References

[1] IBM X-Force Research. 2016. Ransomware: How consumers and businesses value their data. Technical Report (2016)

[2] Gavin O'Gorman and Geoff McDonald. 2012. Ransomware: a growing menace. Symantec Corporation

[3] How Ransomware Became a Billion-Dollar Nightmare for Business. 2016. https://www.theatlantic.com/business/archive/2016/09/ransomware-us/ 498602/. (2016).

[4] Amin Kharaz, Sajjad Arshad, Collin Mulliner, William Robertson, and Engin Kirda. 2016. UNVEIL: A Large-Scale, Automated Approach to Detecting Ransomware. In 25th USENIX Security Symposium (USENIX Security 16). USENIX Association, Austin, TX, 757–772.

[5] Nolen Scaife, Henry Carter, Patrick Traynor, and Kevin RB Butler. 2016. Cryptolock (and drop it): stopping ransomware attacks on user data. In Distributed Computing Systems (ICDCS), 2016 IEEE 36th International Conference on. IEEE, 303–312.

[6] Jiang Ming Dongpeng Xu and Dinghao Wu. 2017. Cryptographic Function Detection in Obfuscated Binaries via Bit-precise Symbolic Loop Mapping. In Proc. 38th IEEE Symposium on Security and Privacy (Oakland'17). San Jose, CA.

[7] Vijayan Prabhakaran, Andrea C Arpaci-Dusseau, and Remzi H Arpaci-Dusseau. 2005. Analysis and Evolution of Journaling File Systems.. In USENIX Annual Technical Conference, General Track. 105–120.

[8] Mendel Rosenblum and John K Ousterhout. 1992. The design and implementation of a log-structured file system. ACM Transactions on Computer Systems (TOCS) 10, 1 (1992), 26–52.

[9] John H Palevich and Martin Taillefer. 2008. Network file system. (Oct. 21 2008). US Patent 7,441,012.

[10] James Gross. 2013. Cloud based storage: A brief look at dropbox. Chronicles 30, 4 (2013).

[11] Understanding Cryptography, A Textbook for Students and Practitioners

[12] Yan Li; Nakul Sanjay Dhotre; Yasuhiro Ohara; Thomas M. Kroeger; Ethan L. Miller; Darrell D. E. Long. "Horus: Fine-Grained Encryption-Based Security for Large-Scale Storage" (PDF). www.ssrc.ucsc.edu. Discussion of encryption weaknesses for petabyte scale datasets.

[13] "The Padding Oracle Attack – why crypto is terrifying". Robert Heaton. Retrieved 2016-12-25.

[14] Researchers crack open unusually advanced malware that hid for 5 years". Ars Technica. Retrieved 2016-12-25.

[15] "New cloud attack takes full control of virtual machines with little effort". Ars Technica. Retrieved 2016-12-25.

[16] Examples of data fragmentation technologies include Tahoe-LAFS and Storj.

[17] Burshteyn, Mike (2016-12-22). "What does 'Active Defense' mean?". CryptoMove. Retrieved 2016-12-25.

[18] CryptoMove Archived 2021-02-06 at the Wayback Machine is the first technology to continuously move, mutate, and re-encrypt ciphertext as a form of data protection.

[19] A. Gazet, "Comparative analysis of various ransomware virii," Journal in Computer Virology, vol. 6, pp. 77-90, 2010..

[20] N. Andronio, S. Zanero, and F. Maggi, "HELDROID: Dissecting and detecting mobile ransomware," in 18th International Symposium on Research in Attacks, Intrusions, and Defenses, RAID 2015 vol. 9404, H. Bos, G. Blanc, and F. Monrose, Eds., ed: Springer Verlag, 2015, pp. 382-404.

[21] A. Bhardwaj, "Ransomware: A rising threat of new age digital extortion," in Online Banking Security Measures and Data Protection, ed: IGI Global, 2016, pp. 189-221.

[22] R. Brewer, "Ransomware attacks: detection, prevention and cure," Network Security, vol. 2016, pp. 5-9, 2016.

[23] The term "cryptovirus" predates ransomware by approximately 10 years. "Cryptovirology" was a term used by Adam Young and Moti Yung in 1996 in their IEEE paper "Cryptovirology: Extortion- Based Security Threats and Countermeasures." Their research and current work can be found at their web site: http: //www.cryptovirology.com/

[24] M. Labs, "Cybercrime tactics and techniques. Q1 2017." Malwarebytes,Tech. Rep., April 2017.

[25] R. Brandom, "Ransomware victims have paid out more than 25 million dollars, Google study finds," July2017, "Last Access: December 19th 2019". [Online].Available: https://www.theverge.com/2017/7/25/16023920/ransomware-statistics-locky-cerber-google-research

[26] T. Ganacharya, "WannaCrypt ransomware worm targets out-of-datesystems," May 2017, "Last Access: December 19th 2019". [Online]. Avail-able: https://www.microsoft.com/security/blog/2017/05/12/wannacrypt-ransomware-worm-targets-out-of-date-systems/?source=mmpc

[27] M. Field, "WannaCry cyber attack cost the NHS£92m as 19,000 appointments cancelled," October 2018,"Last Access: December 19th 2019". [Online]. Avail-able: https://www.telegraph.co.uk/technology/2018/10/11/wannacry-cyber-attack-cost-nhs-92m-19000-appointments-cancelled/

[28] C. Cimpanu, "Ransomware hits hundreds of dentist offices in theUS," August 2019, "Last Access: December 19th 2019". [Online].Available: https://www.zdnet.com/article/ransomware-hits-hundreds-of-dentist-offices-in-the-us/

[29] Jose Miguel Esparza and the Blueliv team, "Spanish consultancyEveris suffers BitPaymer ransomware attack: a brief analysis,"November 2019, "Last Access: December 19th 2019". [Online].Available: https://www.blueliv.com/cyber-security-and-cyber-threat-ntelligence-blog-blueliv/research/everis-bitpaymer-ransomware-attack-analysis-dridex/

[30] "Ransomware hits several Spanish city halls," October2019, "Last Access: December 19th 2019". [Online]. Avail-able: https://www.pandasecurity.com/mediacenter/news/ransomware-spanish-city-halls/

[31] "RSA Security Releases RSA Encryption Algorithm into Public Domain". Archived from the original on June 21, 2007. Retrieved 2010-03-03.

[32] https://pypi.org/project/cryptography/

[33] Machine Learning for Absolute Beginners by Oliver Theobald

[34] Zhang Z, Schwartz S, Wagner L, Miller W. A greedy algorithm for aligning DNA sequences J Comput Biol. 2000 Feb-Apr;7(1-2):203-14.

[35] Machine Learning: An Introduction To Supervised & Unsupervised Learning Algorithms

[36] Lei Yu, Huan Liu Efficient Feature Selection via Analysis of Relevance and Redundancy The Journal of Machine Learning Research Volume 5, 12/1/2004 Pages 1205-1224

[37] https://www.ibm.com/

[38] Machine learning for Beginners, The Definitive Guide to Neural Networks, Random Forests, and Decision Trees by Jennifer Grange

[39] Machine Learning With Random Forests And Decision Trees: A Visual Guide For Beginners by Scott Hartshorn

[40] IBM X-Force Research. 2016. Ransomware: How consumers and businesses value their data. Technical Report (2016)

[41] Gavin O'Gorman and Geoff McDonald. 2012. Ransomware: a growing menace. Symantec Corporation

[42] How Ransomware Became a Billion-Dollar Nightmare for Business. 2016. https://www.theatlantic.com/business/archive/2016/09/ransomwareus/498602/. (2016).

[43] Amin Kharaz, Sajjad Arshad, Collin Mulliner, William Robertson, and Engin Kirda. 2016. UNVEIL: A Large-Scale, Automated Approach to Detecting Ransomware. In 25th USENIX Security Symposium (USENIX Security 16). USENIX Association, Austin, TX, 757–772.

[44] Nolen Scaife, Henry Carter, Patrick Traynor, and Kevin RB Butler. 2016. Cryptolock (and drop it): stopping ransomware attacks on user data. In Distributed Computing Systems (ICDCS), 2016 IEEE 36th International Conference on. IEEE, 303–312.

[45] Jiang Ming Dongpeng Xu and Dinghao Wu. 2017. Cryptographic Function Detection in Obfuscated Binaries via Bit-precise Symbolic Loop Mapping. In Proc. 38th IEEE Symposium on Security and Privacy (Oakland'17). San Jose, CA.

[46] Classification algorithm and its application based on support vector machine by HUANG GANG XIAO HAI JUN

[47] Marcano-Cedeño J. Quintanilla-Domínguez, D. Andina WBCD breast cancer database classification applying artificial metaplasticity neural network Expert Systems with Applications 38 (2011) 9573–9579

[48] Introduction to Support Vector Machines and Kernel-based Machine Learning

[49] Support-vector Machines: History and Applications by Pooja Saigal

[50] I support vector machine

[51] Support Vector Machines, Data Analysis, Machine Learning and Applications, by Brandon H. Boyle

[52] Support Vector Machine, Optimization based Theory, Algorithms, and Extensions by Naiyang Deng, Yingjie Tian, Chunhua Zhang

[53] https://pandas.pydata.org/

[54] https://mode.com/python-tutorial/libraries/pandas/

[55] https://scikit-learn.org/

[56] Support Vector Machines Succinctly by Alexandre KOWALCZYK

[57] One-Hot Encoding in Scikit-Learn with OneHotEncoder, https://datagy.io/sklearn-one-hot-encode/

[58] Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems 2nd Edition, by Aurelien Geron.

[59] Introduction to Machine Learning with Python: A Guide for Data Scientists 1st Edition by Andreas C. Muller Sarah Guido

[60] Artificial Intelligence with Python: A Comprehensive Guide to Building Intelligent Apps for Python Beginners and Developers January 2017 Edition by Prateek joshi

[61] 1. https://docs.microsoft.com/en-us/sharepoint/troubleshoot/security/session-has-expired , Handling ransomware in Sharepoint Online

[62] Kessler, Gary (November 17, 2006).”An Overview of Cryptography”.Princeton University.

[63] Advanced Persistent Security, A Cyberwarfare Approach to Implementing Adaptive Enterprise Protection, Detection, and Reaction Strategies, 1st edition.

[64] Ransom Notes: Know What Ransomware Hit You https://www.trendmicro.com/vinfo/us/security/news/cybercrime-and-digital-threats/ransom-notes-know-what-ransomware-hit-you

[65] Categorical encoding using Label-Encoding and One-Hot-Encoder, https://towardsdatascience.com/categorical-encoding-using-label-encoding-and-one-hot-encoder-911ef77fb5bd

[66] Pandas: Python Data Analysis Library, https://medium.com/nerd-for-tech/pandas-python-data-analysis-library-1d061c982fc8

[67] Python Pandas by ayush shekhar https://medium.com/analytics-vidhya/python-pandas-part-1-7f4065283982

[68] Python Machine Learning 3 Books in 1 - the Ultimate Beginners, Intermediate and Expert Guide to Master Python Machine Learning