A Survey on Detection & Prevention Methods of Ransomware in Cyber Security

1Malhar Pandya, 2Snehal Sathwara
1Student, 2Assistant Professor
R K University

Abstract - **This paper examines the contributions of research to ransomware malware detection that uses machine learning and deep learning algorithms. The main motivations for this research is the destructive nature of ransomware, the difficulty of reversing the infection of ransomware, and the importance of detecting it before infecting the system. Machine learning soon coming at the forefront of the fight against ransomware, so here, we have tried to identify the weak points of the machine learning approaches and how they can be reinforced.**

keywords - **Malware, Ransomware, Cyber Security, Machine Learning**

I. INTRODUCTION

Nowadays, the use of internet, computers and smart devices is very common and a lot of people use these devices on a daily basis. As we go around there are people with good and bad intentions, this also happens in the internet world because there are people with bad motives who want to take advantage of real users for their own personal gain. To combat the threat of advanced versions of malware, new detection mechanisms are needed. Possible solutions to this problem are that signature-based analysis can be integrated with machine learning techniques that can provide high accuracy compared to a signature-based approach established used for detection. Machine learning algorithms can be used to look at these links and potentially detect malware more accurately.

Malware is a software that is intended to destroy or disrupt computers and computer systems. It is a short for "malicious software." Viruses, worms, Trojan horses, spyware, adware, and ransomware are examples of common malware. Ransomware (and other malware) is frequently spread through email spam campaigns or targeted attacks. To establish its presence on an endpoint, malware requires an attack vector. After establishing its presence, malware remains on the system until its mission is completed.

II. LITERATURE SURVEY

1) Analyzing Ransomware using Machine Learning

Ransomware attacks have become wild these days in various areas like schooling, health & wellbeing, business, research, furthermore, data innovation and information technology. These attacks are a consequence of different social engineering methods used to appeal a user to click on a vindictive & malicious attachment, links, etc. in an email or through alternate ways.

The attacks can occur because of its triggering, execution properties and propagation. It can likewise use cryptographic methods to lock the system, utilize the crypto-currencies and remote control and command channel. Besides, ransomware attacks take advantage of system's shortcomings like Windows Server Message Block (SMB), Remote Code Execution Vulnerability, to get in and lock the system.

Ransomware can be classified into following two classes [1]:

i. Crypto ransomware
   Crypto ransomware encrypts the documents in user's system, making the records and files inconceivable to use except if it is decrypted. Eliminating the ransomware or taking the hard drive to an unaffected system doesn't address the issue as the user doesn't have the decryption key to access the data. The user is approached to pay the ransom explicitly in the type of crypto-currency to decrypt and recover the initial original records.
   Bitcoin is broadly utilized by the attackers because of its namelessness as the character of the attacker is hard to track. Paying the ransom amount doesn't ensure that the user will get the key to decrypt and recover the records and data. A few instances of Crypto ransomware are: Locky, CryptoLocker, CryptoWall and SamSam.
ii. Locker ransomware
   Locker ransomware locks the user's system because of which the system becomes distant and inaccessible. Nonetheless, unlike Crypto ransomware, the user's documents and records are not encoded. Here, the attacker might ask ransom in form of as amount vouchers. The user may move the hard drive to an unaffected system and gain admittance to the locked documents. Instances of Locker ransomware include: Winlocker and CTB-storage.

2) Ransomware Detection Process for learning algorithms

Location of ransomware, utilizing machine learning or deep learning follows a quite pattern, as portrayed in Figure 1. There should be selection of features, be it custom feature IoT 2020, 1 555 selection strategies or algorithms [2]; when this is finished, and the ideal list of feature set is found, the data, coordinated by these features, will be taken care of into the choice of machine learning algorithm or deep learning algorithm. The algorithm will be prepared and afterward tested. Learning algorithms permit the system to learn without anyone else supervising. Learning algorithm work in various ways, the principal ways being supervised and unsupervised learning.

Algorithms of Supervised Learning will require a preparation or training set, for the case of ransomware data of both harmless and ransomware is required, so the algorithm can figure out how to distinguish patterns that separate the two from one another. Unsupervised Learning techniques take care of datasets that are not named and will try to find patterns which can create models, to recognize the data types.

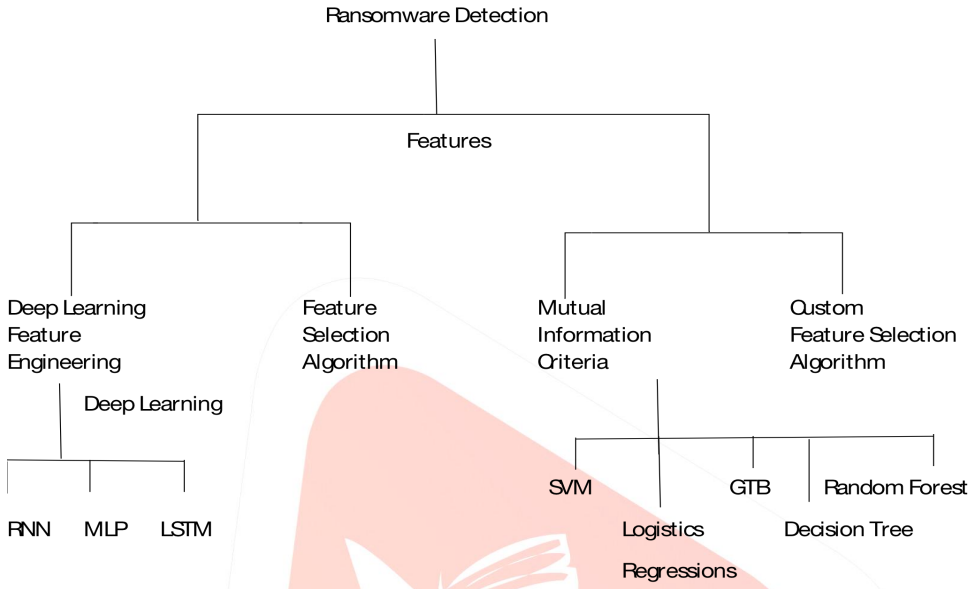

Figure 1: Ransomware detection taxonomy

3) Ransomware Encryption Process

Encryption is an important element of the ransomware business model, as the success of attack ransomware depends on whether the victim's files are effectively encrypted. Symmetric encryption has the advantage of the speed the attacker wants, as the attacker does so much encryption. It tries to encrypt as many files as possible but has a problem with key management.

Asymmetric encryption overcomes the challenges of key management and improves restoring force. The ransomware attacker has deployed hybrid encryption that uses the symmetric key (Ksecret) to encrypt the victim's file and the public key (Kp) to encrypt the Ksecret. In this way, symmetric key can be decrypted only with the private key Ks from the generated RSA pair. Some attack implementations store the encrypted symmetric key on the victim's system, while others steal it to C & C, such as the infected chain shown in Figure 2 below. The chain begins with an attacker generating RSA key pairs Kp and Ks and embedding only the public key Kp. Put it in the ransomware payload for the first infection. The payload produces a symmetric key Ksecret and / or beacon to the C & C.

Ransomware encrypts target files mi it uses the symmetric key Ksecret to offload the corresponding cipher text Ci, and immediately after encryption of the target file is complete, encrypts the Ksecret and offloads another cipher text Cj. The symmetric encryption algorithm is used by the dependents on the ransomware family. This is because others are known to use AES [12], while other algorithms use custom ones. The Ksecret cipher text Cj is stored on the victim's machine or compromised by C & C. The victim will eventually receive a ransom note notification. There is no guarantee that Ks will be provided with the key K for decryption from Ks secret, as the attacker could continue to exploit the victim.

4) Infection Vectors in Ransomware

The structure of the ransomware attack follows the method shown in Figure 2. Different Infection vectors usually carry out ransomware infections. The first is the most prominent vector Malicious emails and payloads are delivered as email attachments from emails sent via spam Botnets and other compromised hosts [5]. Exploit kits are another well-known method of infection. Exploit kits are software packages that scan your system for vulnerabilities in order to infect it. Use malware [4]. Another well-known method of infection is drive-by download. The victim, will be directed to the malicious Web site to execute malicious code. The installation takes place after the payload has been placed on the system. A Prominent method the installation is a download dropper. This approach uses an initial file that you must use small code to avoid detection and reach the Command-and-Control Center (C & C). Ransomware of the author, in order to avoid the AV, tries to split the execution in various scripts and processes. To avoid (Antivirus) Signature-based detection. Targeted ransomware if your organization is under attack Networks spread by locating file shares, infecting them and maximizing

them. Increases confusion and potential ransom. The executable will not run till multiple machines are infected one by one.

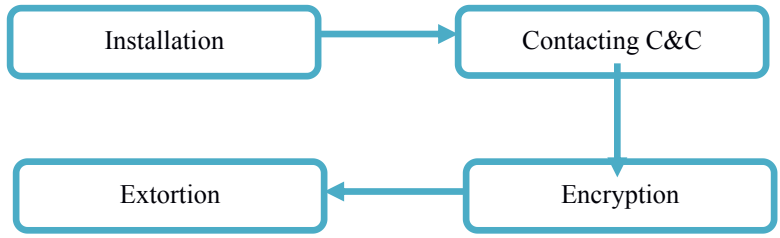

Figure 2: Structure of Ransomware Attack

5) Random Forest Algorithm in Machine Learning

The Random Forest algorithm [3] is a classification algorithm. It is an easy-to-use, flexible, and one of the most commonly used machine learning algorithms. This is because it is simple and can be used for both classification and regression tasks. Random Forest is a supervised learning algorithm. You can tell by the name, it creates a forest and randomize it. Forest Built is an ensemble of Decision Trees. In the Random Forest algorithm, the trees are built from the specified set of attributes, and each tree gives a result. The one that results in a majority is voted as the final output. Therefore, there are chances that random forest gives the best results compared to others. There are many advantages and disadvantages to using the Random Forest algorithm.

Advantages of Random Forest Algorithm:

i. Useful for both classification and regression.
ii. It can process missing values and maintain the accuracy of missing data. Not compatible with models.

Disadvantages of Random Forest Algorithm:

i. Classification works well, but not as good for regression.
ii. One can hardly control the functions of model.

This figure describes the workflow of the Random Forest algorithm. This shows the workflow of the Random Forest algorithm. It mainly consists of 6-7 levels. It's very easy to understand and use the Random Forest algorithm. Therefore, it is used in many cases in training malware detection models for the highest accuracy.

Random Forest Workflow (shown in Figure 3) [3]:
Step 1: The specified attribute should be M and the number of trees in the forest should be n (estimator's number in forest).
Step 2: Randomly select the m attribute from the specified M attributes.
Step 3: Build a tree with m selected attributes.
Step 4: Repeat steps 2 and 3 n times.
Step 5: From a collective decision, each tree finds the resulting maximum iteration result.
Step-6: Until the tree is build, repeat the steps 1 to 5

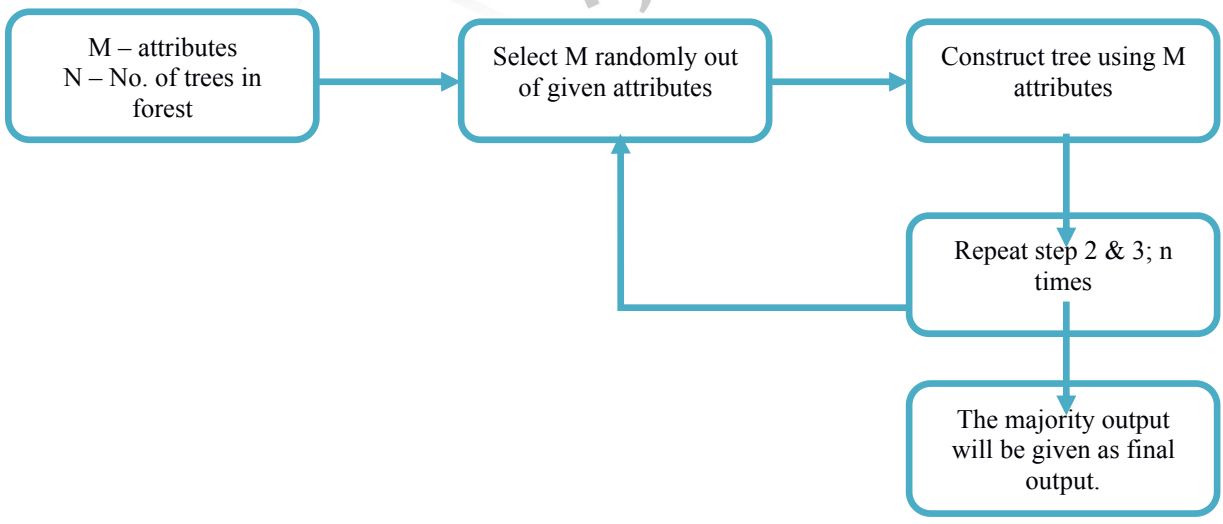

Figure 3: Ransomware detection taxonomy

CONCLUSION

Almost all attacks are summarized in a simpler approach. Phishing or spam is often used to gain access to end-user devices to carry out attacks, enabling simple, high-level malware installations. A good example of this is a Trojan horse that needs to be attached to a file. Protection from this type of software relies on appropriate instructions to the end user. For example, do not click unknown links or buttons, or download files from unknown sources. If your computer contains malicious files, these instructions will not trigger the virus. Of course, some attacks require much more complex protection, but with appropriate safeguards such as firewalls, antivirus software, and employee training, you can stop the breach, or at least minimize the damage. I can. Overall, these measures can significantly reduce the potential for damage, but because biological viruses occur in the same way as nature, and humans (hackers) modify and develop computer viruses. It can never be zero. Therefore, full protection is not possible. However, if each person pays a little more attention to the protection of devices such as computers, phones and tablets, in combination with the efforts of network specialists, higher protection will be achieved, resulting in attacks. This article presents the latest developments in cybersecurity issues and some recommendations on how to deal with or prevent cyberattacks. Developing network and information security recommendations and guidelines for universities, research institutes, schools, and government is the basis of much larger research.

REFERENCES

[1] Fernando, Damien Warren, et al. "A Study on the Evolution of Ransomware Detection Using Machine Learning and Deep Learning Techniques." IoT, vol. 1, no. 2, 2020, pp. 551–604., https://doi.org/10.3390/iot1020030.

[2] Zimba, A. Malware-Free Intrusion: A Novel Approach to Ransomware Infection Vectors. Int. J. Comput. Sci. Inform. Secur. 2017, 15, 317–325.

[3] Ganta, Venkata Gopi, et al. "Ransomware Detection in Executable Files Using Machine Learning." 2020 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT), 2020, https://doi.org/10.1109/rteict49044.2020.9315672.

[4] Noorbehbahani, Fakhroddin, and Mohammad Saberi. "Ransomware Detection with Semi-Supervised Learning." 2020 10th International Conference on Computer and Knowledge Engineering (ICCKE), 2020, https://doi.org/10.1109/iccke50421.2020.9303689.

[5] CyberPedia. What Is an Exploit Kit. 2018. Available online: Https://www.paloaltonetworks.com/ cyberpedia/what-is-an-exploit-kit (accessed on 21 November 2018).

I.