Trusted Detection of Ransom ware using Machine Learning Algorithms

Shemitha P.A, Julia Punitha Malar Dhas

Abstract*— Nowadays, the Computer Networks and the internet are increased. Lots of information is accessed and allowed to the users to share the information to the Internet. One of the major issues with internet was different types of attack. Ransomware is a one kind of attack or it is malicious software that threatens to publish the victim's data. A variety of threats is the main target for the effective network security and avoids them from spreading or entering to the networks the network security on computer essential for computer networks. Ransom ware is a critical threat in network security since each day the raising of ransomware gets abundant. The major problem by the researchers is the prediction of ransomware. This paper planned to carry out a review on the different method to detect ransomware. Ransomware detection is very much helpful on minimizing the workload of analyst and for determining the variation in hidden Ransomware samples. Using machine learning algorithms Ransomware detected efficiently and trustfully.

INTRODUCTION

Ransomware is type of malicious software. Based on restricting user access to their own data ransomware extract ransom from users. Encryption is the most popular way to restrict access to user data, and thus, ransomware that uses encryption techniques is called Crypto ransomware. Ransomware uses sophisticated public-key cryptography that is impossible to crack and goes for certain types of files that is supposed to be the most valuable to the user, such as text documents, images and specialized formats.

One most popular type of ransomware is encryption based ransomware. Other types are locker ransomware, scareware and fake ransomware Locker ransomware is designed to block user’s access to the system or specific applications. Ransomware like that either replaces the desktop with a custom one, or targets popular apps like browsers by modifying certain files. Scareware is one type of ransomware that uses scare tactics in order to force users into paying ransom. It also uses social engineering and other tactics in order to make users pay. One of the most common tactics is to display a message from law enforcement that includes personal information such as location and name of the ISP provider, making the message more believable. The message will demand a “fine” for certain made-up offense, such as copyright infringement or watching child pornography, and threatens user that if they refuse to pay they will be jailed. Many modern ransomware don’t bother encrypting user’s data. Instead, they just delete it right away, creating a bunch of dummy files in order to fool the user into thinking that their data is still recoverable called Fake ransomware. To distinguish fake and paid ransomware is impossible, it is always best to never pay ransom, unless the situation is actually critical.

There are different ways that ransom ware uses to get into your system. Using infected spam emails is the most popular method, that are usually distributed by vast networks of botnets. Such an email will usually contain the message, that message uses social engineering techniques in order to prompt the user to click on the link to be infected or malicious attachment will be downloaded. Another method of spreading ransom ware is infected adverts on the net. Once the user clicks on the advert, a JawaScripts starts running that javaScript is malicious. Then downloading a payload on the user’s PC. Beyond that, ransom ware can be spread on removable drives, or it self-propagate via a network by searching for open ports and unprotected connections. Perpetrators will also use exploit kits in order to leverage vulnerabilities that unknown and get ransom ware into your system. Once there, it will phone back (usually, without encrypting network traffic), and then looking for certain types of data to encrypt. After the encryption of data, a ransom note is displayed. Ransomware uses various techniques in order to protect yourself from being detected or analyzed, including obfuscation and system mapping and it is designed to distinguish between real system and a honey pot.

RELATED WORKS

Guan *et al.* [1] have proposed a GRBC that was adopted as a metric and hence a successive algorithm was introduced to calculate this GRBC. SDS key group of nodes in underlying network was found by the GRBC that in which the security devices and NSFs were installed. The outcomes have shown the improvement over the security performance in SDS systems. Dodangeh and Jahangir [2] have developed a new scheme to satisfy the security in WBAN. Two mutual authentication and key exchange protocols were proposed to deal with the overall network architecture in WBAN circumstances. BAN logic was used for verification, and the result has shown that all the WBANs communications were solved from the medical server biometrics.

Li *et al.* [3] has modeled the airport security checkpoints by various passenger approach, demonstrates possible network structures and the various performance was compared in the case of queuing. The findings shown that M/M/1 systems combination has a same or better performance over the M/ M/n system when considering the

Revised Manuscript Received on July 18, 2019.

Shemitha P.A, Research Scholar, Computer Science & Engineering, Noorul Islam centre for higher education, Kannyakumari.

Dr. Julia Punitha Malar Dhas, Professor &Head of the Department, Computer Science & Engineering, Noorul Islam centre for higher Education, Kannyakumari.

passengers' strategies and feelings. In 2018, Sharma *et al.* [4] have developed a narrative authentication protocol and key exchange in which Xhaul links was secured by the movement terminal of network. The analysis thus satisfied the performance evaluation and security requisites effectively using the AVISPA tool and BAN logic than the conventional models. In 2018, Hyun *et al.* [5] have presented the development and modeling of the architecture called I2NSF. Here the SDN-integrated I2NSF architecture was implemented along with its applications of security. Further, the research challenges and numerous standardization of I2NSF were discussed.

In 2018, Shi *et al.* [6] have made a study for enhancing the physical layer security. Here, the relays were collected as various clusters to improve the possibility of secrecy outage. The optimization and the performance were offered by two methods named: secrecy outage probability and average secrecy capacity and the results were analyzed. In 2018, Tubail *et al.* [7] have implanted an algorithm to design and transmit minimum power preserving data streams. The joint optimization problem was solved by the iterative optimization algorithm and SDP. Four transmission models were implemented, and the experimental results have shown a better efficiency in transmission security over the IA based multiuser relaying networks. In 2017, Kraus *et al.* [8] have focused in intrinsic motivation analysis on fulfilment of psychological need. The security and the other needs were the major one to employ the privacy and security actions on the Smartphone. Beyond the security needs, the addressing psychological needs were an important one to be focused.

In 2017, Tang *et al.* [9] have enhanced the physical layer security by full-duplex users for heterogeneous networks to downlink. To solve the jammer selection threshold, a greedy algorithm was proposed. The theoretical analysis accuracy and efficiency were estimated in the experimentation. In 2015, Pourazarm and Cassandras [10] have studied the maximizing problem with respect to the initial energy allocation and routing over the sensor network nodes. The observation was done on the network’s performance under security threats that were made by attacks on faked-cost. Thereby they made a robust optimal routing probabilities and network lifetime under the routing attacks. In 2016, Kaynar [11] have introduced a classified structure for learning by the attack graph generation process of the method that was given in every phase. Based on the proposed classification structure, the literature related works were stated, and the possible challenges and open issues were made by the contributive ideas.

In 2016, Hu *et al.* [12] have introduced a strategy with security protection to transmit adaption mechanism. Mainly, a CJ scheme replaces cooperative transmission, when there was no satisfaction in security or QoS constraint. The modeling was better in secured transmission, flexible and efficient in power resource utilization. In 2018, Ahlawat and Dave [13] have investigated the node capture attack issues and proposed a safe HKP-HD for WSN. The objective of the proposed method was to provide more resistant against the node capture attacks for the network. When comparing with the conventional method, the proposed model possessed a minimum probability of communication overhead, storage overhead and of key compromise.

In 2017, Meng *et al.* [14] have presented the trust-based approach by utilizing Bayesian inference for identifying the malicious nodes in MSN circumstances. The efficiency was demonstrated by evaluating the proposed approach in a real-world environment for detecting malicious nodes with two healthcare groups. In 2011, Wen *et al.* [15] have introduced a channel identification characteristic for detecting the node clone attack in WSN. With regards to the traditional models, the implemented model provided a perfect flexibility of node clone attack with certain parameters like high detection probability, low memory requirements, etc. In 2018, Kumar and Palanichamy [16] have proposed an S-SELDRIP in which the hop by hop authentication model has offered a secured optimal routing while data dissemination. The outcome was obtained by testing the proposed model in NS2 Mannasim framework, and it shown a better viability.

RANSOMWARE DETECTION & RESULTS

The first variants of Ransomware used some file extensions like .crypt. However, each new variant use different extensions and some of them the filename keep intact. Because of this, the multiple symptoms of an attack need to be watch. Here take five of them.

A. Watch out for known file extensions

The number of known Ransomware file extensions is growing very fast; it is a useful method for detecting malicious activity. User must need to get file activity monitoring in place before the user do anything, so that you have both a real time and historical record of all folder activity and file on your network file shares.

B. Watch out for an increase in file renames

In the activity on network file shares file renames are not a common action. In normal case, user may end up with just a handful of renames even if your network have more than hundreds of users. When Ransomware strikes, it create more increase in file renames as your data gets encrypted. An alert is triggered using this behavior. A threshold value was set and the total number of rename is more than this thresholds, then user have potential ransomware effect.

C. Create a Sacrificial Network Share

If the ransomware strikes, it first looks for local files and move to network shares. Most of them are go through network shares in alphabetical order. A sacrificial network share is a early warning system ans also delay ransomware from our critical data.

D. Update your IDS systems with exploit kit detection rules

Many IDS, IPS and firewall systems are come with detection features. The methods to get Ransomware onto a client via malspam or via compromised websites are Exploit kits. Most common exploit kits are associated with Ransomware is Neutrino EK and the Angler EK. Check if your network security monitoring systems and also check if they have the capability to detect exploit kits.

E. Use client based anti-ransom ware agents

Recently some of the companies released anti-ransom ware software applications. These applications are run in the background and block Ransomware to encrypt data. They also monitor the Windows registry for text strings known to be associated with Ransomware. In this approach the users need to install client software on every network device. Researchers are also looking at ways to ‘crash’ computer systems when droppers are detected. Droppers are small applications that first infect target machines. Droppers are infects machine in preparation for downloading the main malware payloads. This will likely mean that the system is sent to IT where the attack should be discovered.

PROPOSED METHOD

Crypto ransom ware is considered as one malware which blocks the user file’s access by encrypting them and in turn demands them with ransom for attaining the decryption key. Hence, this poses a serious warning over the most of the companies. Hence, ransom ware detection is very much helpful on minimizing the workload of analyst and for determining the variation in hidden ransom ware samples. A new ransom ware detection method will be proposed with the intension of attaining the better detection rate. The new ransom ware detection model will be implemented using the machine learning approach. The proposed detection model includes three stages: (i) Feature extraction, (ii) Feature selection and (iii) classification. In feature extraction process, the sequential pattern features are extracted. . After the feature selection process, these optimal sequential patterns will be subjected to do the classification process.

A. Feature extraction

Machine learning algorithms learn or detect the specific patterns from feature vectors (feature vector is fixed length), and therefore feature extraction is the first step before using the classification algorithm for Ransomware detection. For features that are in the form of sequences, such as sequences of code bytes, operation codes (opcodes), system calls, or API calls, the creation of a representative feature vector is a nontrivial problem. There are different feature extraction methods such as binary feature extraction, frequency feature extraction, frequency weight feature extraction, hidden markov model etc. Binary feature extraction is the basic feature extraction of feature from sequence and it identifies the all the distinct element from the feature. The sequence can then be represented as a binary vector. Frequency feature extraction method, the count of occurrence of a dictionary term in the sequence is used instead of just its presence or absence. Frequency weighting methods such as term frequency-inverse document frequency (TF-IDF) have also been employed to generate feature vectors from sequences. An HMM represents a doubly stochastic system in the form of a finite state in which the states are not visible. The processing or progress of the state machine is observed through certain symbols. The symbols are emitted in each state. The state machine employ two type of probabilities: the state transition probabilities determine the next state for any given state, and symbol emission probabilities indicate the possibility of each symbol being emitted in a given state.

B. Feature selection

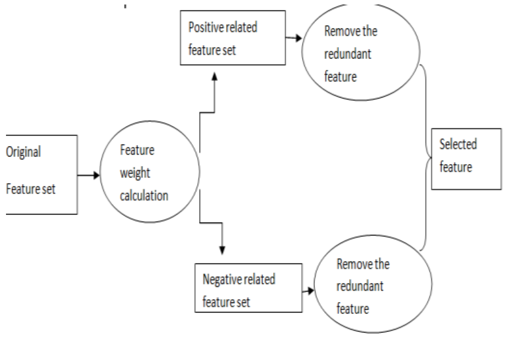
Positive related feature set
Remove the redundant feature
Original Feature set
Feature weight calculation
Selected feature
Negative related feature set
Remove the redundant feature

Fig1: Feature selection

In the feature selection stage feature weight calculated from original feature set. Then the feature sets are grouped into two such as positive related feature set and negative related feature set. From these two feature sets redundant features are eliminated and new feature set created this is used for classification.

C. Classification

The Ransomware detect using neural network, support vector machine and random forest method, these three method used as classifier in Ransomware detection. The accuracy of neural network is 95%, the accuracy of support vector machine is 78% and the accuracy of random forest is 80%. So the proposed system use the random forest algorithm because it’s accuracy is better than other two algorithms. The random forest classifier is implemented in Python due to the random forest’s low computational training cost and its use of bagging to reduce variance and overfitting. A random forest classifier is an ensemble algorithm, which utilizes a collection of decision trees to vote and predict the class of the input data. Using the random subset of the feature set each decision tree is created. Each decision tree is generated using the gini impurity metric, which measures the probability of mislabeling a randomly chosen element from the training set if the element was labeled based solely on the distribution of the binary labels in the set. Three of the main tuning metrics for a random forest classifier include the number of decision trees in the forest, the depth of each decision tree, and the maximum number of features that can be included in each decision tree. The number of trees in the forest dictate the performance and variance of the classifier. A larger number of trees results in higher classification accuracy and lower variance but increases the computational cost of the classifier. The depth of each tree has a similar cost-benefit situation. As the depth of each tree increases, the induced bias in the classifier decreases; however, the added depth comes with a computational penalty. The last main metric we used for tuning our random forest classifier is the maximum number of features that can be included in each decision tree. The maximum number of features is used to determine the best split when creating a decision tree. Once again, increasing the number of features increases performance but comes at a computational cost.

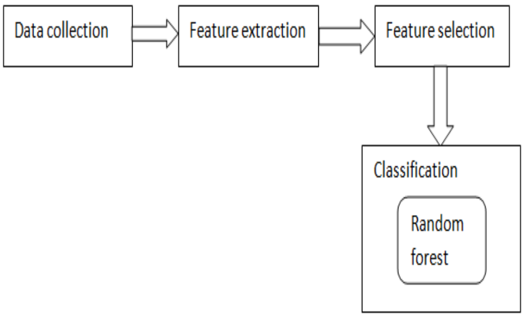

Fig2: machine learning approach to detect Ransomware

CONCLUSION

In this paper we reviewed different techniques for ransom ware detection. Many techniques have been used in detection of ransom ware, but machine learning techniques are common in recent literatures. Due to higher accuracy and low computational cost we selected random forest classifier algorithm for ransom ware detection.

REFERENCE

1. AaronZimba,ZhaoshunWang,HongsongChen," Multi-stage crypto ransomware attacks: A new emerging cyber threat to critical infrastructure and industrial control systems", ICT Express, vol.4, no.1, pp.14-18, March 2018
2. S. Homayoun, A. Dehghantanha, M. Ahmadzadeh, S. Hashemi and R. Khayami, "Know Abnormal, Find Evil: Frequent Pattern Mining for Ransomware Threat Hunting and Intelligence," IEEE Transactions on Emerging Topics in Computing, 26 September 2017.
3. DanielMorato, EduardoBerrueta, EduardoMagaña, MikelIzal," Ransomware early detection by the analysis of file sharing traffic", Journal of Network and Computer Applications, vol.124, pp.14-32, 15 December 2018
4. HanqiZhang, XiXiao, FrancescoMercaldo, ShiguangNi, FabioMartinelli, Arun KumarSangaiah," Classification of ransomware families with machine learning based on N-gram of opcodes", Future Generation Computer Systems, vol.90, pp.211-221, January 2019
5. D. Min et al., "Amoeba: An Autonomous Backup and Recovery SSD for Ransomware Attack Defense," in IEEE Computer Architecture Letters, vol. 17, no. 2, pp. 245-248, 1 July-Dec. 2018.
6. L. J. García Villalba, A. L. Sandoval Orozco, A. López Vivar, E. A. Armas Vega and T. Kim, "Ransomware Automatic Data Acquisition Tool," in IEEE Access, vol. 6, pp. 55043-55052, 2018.
7. Alfredo Cuzzocrea, Fabio Martinelli, Francesco Mercaldo, Giorgio Mario Grasso," Experimenting and assessing machine learning tools for detecting and analyzing malicious behaviors in complex environments", Journal of Reliable Intelligent Environments, vol.4, no.4, pp 225–245, December 2018
8. Aniello Cimitile, Francesco Mercaldo,Vittoria Nardone, Antonella Santone, Corrado Aaron Visaggio," Talos: no more ransomware victims with formal methods", International Journal of Information Security, vol.17, no.6, pp 719–738, November 2018
9. ZhaoDongmei, LiuJinxing," Study on network security situation awareness based on particle swarm optimization algorithm", Computers & Industrial Engineering, vol.125, pp.764-775, November 2018
10. JianfengGuan, ZhijunWei, IlsunYou," GRBC-based Network Security Functions placement scheme in SDS for 5G security", Journal of Network and Computer Applications, vol.114, pp.48-56, 15 July 2018
11. AbdussalamSalama, RezaSaatchi," Probabilistic classification of quality of service in wireless computer networks", ICT Express, Available online 5 October 2018
12. NicolaAccettura, GiovanniNeglia, Luigi AlfredoGrieco," The Capture-Recapture approach for population estimation in computer networks", Computer Networks, vol.89, pp.107-122, 4 October 2015
13. StearnsBroadhead," The contemporary cybercrime ecosystem: A multi-disciplinary overview of the state of affairs and developments", Computer Law & Security Review, vol.34, no.6, pp.1180-1196, December 2018
14. LucaTosoni," Rethinking Privacy in the Council of Europe's Convention on Cybercrime", Computer Law & Security Review, vol.34, no.6, pp.1197-1214, December 2018
15. KrzysztofCabaj, MarcinGregorczyk, WojciechMazurczyk," Software-defined networking-based crypto ransomware detection using HTTP traffic characteristics", Computers & Electrical Engineering, vol.66, pp.353-368, February 2018
16. Jane Y.ZhaoMD, MS, Evan G.KesslerMD, JihnheeYuPhD, KabirJalalPhD, Clairice A.CooperMD, FACS, Jeffrey J.BrewerMD, FACS, Steven D.SchwaitzbergMD, MA, FACS, Weidun AlanGuoMD, PhD, FACS," Impact of Trauma Hospital Ransomware Attack on Surgical Residency Training", Journal of Surgical Research, vol.232, pp.389-397, December 2018