Ransomware Attack Modeling and Artificial Intelligence-Based Ransomware Detection for Digital Substations

Syed R. B. Alvee
Dept. of Electrical Engineering and Computer Science
Texas A&M University-Kingsville
Kingsville, TX 78363 USA
syed_raqueed_bin.alvee@students.tamuk.edu

Bohyun Ahn
Dept. of Electrical Engineering and Computer Science
Texas A&M University-Kingsville
Kingsville, TX 78363 USA
bohyun.ahn@tamuk.edu

Taesic Kim*
Dept. of Electrical Engineering and Computer Science
Texas A&M University-Kingsville
Kingsville, TX 78363 USA
seerin.ahmad@students.tamuk.edu

Ying Su
Dept. of Computer Science
The University of Texas at Austin
Austin, TX 78712 USA
ys24487@utexas.edu

Young-Woo Youn
Adv. Power Apparatus Researh Center
Korea Electrotechnology Research Institute
Changwon, 51543 South Korea
ywyoun@keri.re.kr

Myung-Hyo Ryu
Adv. Power Apparatus Research Center
Korea Electrotechnology Research Institute
Changwon, 51543 South Korea
mhryu@keri.re.kr

2021 6th IEEE Workshop on the Electronic Grid (eGRID) | 978-1-6654-4979-3/21/$31.00 ©2021 IEEE | DOI: 10.1109/EGRID52793.2021.9662158

Abstract—**Ransomware has become a serious threat to the current computing world, requiring immediate attention to prevent it. Ransomware attacks can also have disruptive impacts on operation of smart grids including digital substations. This paper provides a ransomware attack modeling method targeting disruptive operation of a digital substation and investigates an artificial intelligence (AI)-based ransomware detection approach. The proposed ransomware file detection model is designed by a convolutional neural network (CNN) using 2-D grayscale image files converted from binary files. The experimental results show that the proposed method achieves 96.22% of ransomware detection accuracy.**

Keywords—artificial intelligence, attack modeling, convolutional neural network, cybersecurity, digital substation, ransomware

I. Introduction

Ransomware is a malware which usually encrypts the important or credential data and theft of control of a system in demand of a ransom. A ransomware attack was discovered early in September 2013 [1]. The attack initiated by sending emails with an attachment that contains CyrptoLocker ransomware (i.e., an executable file (.exe), but disguised as a normal PDF file). The executed ransomware encrypted certain types of files on local hard drives and network drives using Rivest–Shamir–Adleman (RSA) public key cryptography, while the malware control servers stored the private key which was only provided if a payment was made. It is also observed that the ransomware file was able to be propagated using Gameover Zeus trojan and botnet as well [1]. The ransomware attacks became a global concern after more than 1,400,000 Kaspersky users were attacked across various sectors in 2016 [2]. In 2017, about 400,000 machines in 150 countries were infected by the WannaCry ransomware [3]. Therefore, many security researchers in information and communications technology (ICT) domains have paid special attention to ransomware detection in recent years.

Recently, ransomware attacks have targeted industry control systems (ICS) and increased about 500% from 2018 to 2020 [4]. In 2021, the Colonial Pipeline suffered a ransomware attack that impacted computerized equipment managing the pipeline. The company provided 4.4 million dollars in Bitcoin for the decryption tool [5]. It is anticipated that more and more ransomware attackers will target critical power infrastructures such as substations and wind/solar farms. Therefore, it is crucial to early detect and mitigate ransomware attack.

In general, ransomware detection methods are classified into two categories: Static analysis and dynamic analysis methods. Static malware analysis examines ransomware without executing the actual binary files. Simple static analysis methods utilize static data such as file header information, file hash, and URL. There are open-sources tools/servers providing static malware analysis such as VirusTotal [6]). Although conventional static malware analysis methods are simple to detect known malware and easy to implement [7], they are largely ineffective against sophisticated ransomware attacks [8]. Recently, static malware detection methods using artificial intelligence (AI) have been proposed to improve detection accuracy [9]. Since ransomware is evolving, new malware should be detected as well.

Dynamic analysis methods detect ransomware attacks using abnormal behavioral data caused by the compiled ransomware or ransomware events by adversaries in the target system. The authors in [10] collected packets and data from network traffic between an infected computer and a command and control (C2) server. Using the network data, a random forest (RF) machine learning (ML) method detected ransomware with over 86% of detection accuracy. In [11], a ML combining Navies Bayes (NB) and Support Vector Machine (SVM) is used to detect ransomware attacks by using network data from function virtualization (NFV) and software defined network (SDN). The authors claim that this approach could achieved 99.99% detection rate. Comparing to static malware analysis methods, dynamic methods might provide better capability of detecting sophisticated and unknown ransomware. However, a huge amount of network and event data are required.

This work was supported in part by the Korea Electrotechnology Research Institute (KERI), the National Science Foundation (NSF) under award No. EEC-1359414, and the Department of Energy (DOE) under award No. DE-EE0009026.

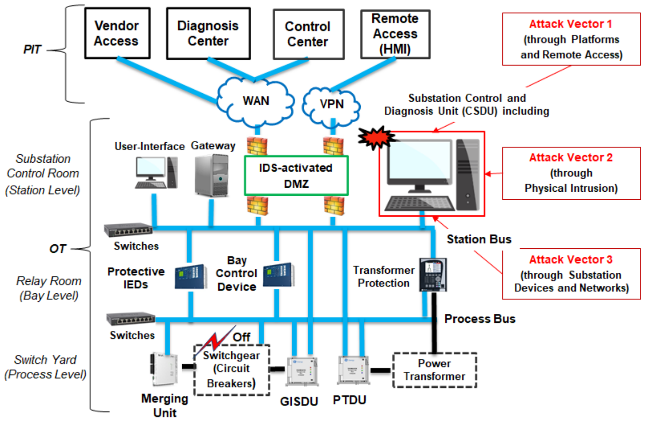

PIT
Vendor Access
Diagnosis Center
Control Center
Remote Access (HMI)
Attack Vector 1 (through Platforms and Remote Access)
WAN
VPN
Substation Control and Diagnosis Unit (CSDU) including
Substation Control Room (Station Level)
User-Interface
Gateway
IDS-activated DMZ
Attack Vector 2 (through Physical Intrusion)
Switches
OT
Station Bus
Relay Room (Bay Level)
Protective IEDs
Bay Control Device
Transformer Protection
Attack Vector 3 (through Substation Devices and Networks)
Switches
Process Bus
Off
Switch Yard (Process Level)
Switchgear (Circuit Breakers)
Power Transformer
Merging Unit
GISDU
PTDU

Fig. 1. Ransomware attack vectors targeting a local server in a digital substation.

The goal of this paper is to explore ransomware attacks in a digital substation and investigates an artificial intelligence (AI)-based ransomware detection method. A cyber kill chain (CKC)-based ransomware attack modeling method is designed, which targets disruptive operation of a digital substation. A convolutional neural network (CNN) model is designed and trained using 2-dimesional (2-D) grayscale image files from real ransomware files. Experimental results validate the proposed CNN-based method with good detection results.

II. Ransomware Attack Modeling in Sub-station

Fig. 1 introduces three potential ransomware attack vectors, targeting to disrupt a substation control and diagnosis unit (CSDU) operation in a local substation control room. An attacker's goal is to encrypt the CSDU local system by ransomware. Attack vector 1 is an external network attack path initiated from the platform information technology (PIT) involving vendor access servers, diagnosis centers, control centers, and other remote access points. Attack vector 2 is a local network attack route started from the internal substation in the operational technology (OT). Attack vector 3 is a physical intrusion. An intrusion detection system (IDS)-activated DMZ is established between the PIT and OT networks. The IDS implements a deep packet inspection and ransomware detection programs against all incoming ransomware from the PIT network.

The CKC model is an attack modeling method that describes the chain of a cyber threat actor's actions in terms of attack tactics, techniques, and procedures. The latest substation related CKC version is MITRE's ATT&CK for ICS framework [12] which enumerates the actions of a cyber adversary might occur with an ICS environment. This paper design attack models on a digital substation based on the MITRE's ATT&CK for ICS framework. Fig. 2 shows a CKC ransomware attack model for a digital substation having twelve ransomware attack phases. The attack scenario has been created by referring the Colonial Pipeline ransomware attack incident reports [13]. An advanced persistent threat (APT) actor is accessed a PIT system (e.g., a control/diagnosis center server) by social engineering (e.g., phishing) or exploiting remote access accounts leaked in the dark web (1. Initial Access). A backdoor malware is established then executed in the system (2. Execution). The adversary is continuously maintaining a foothold using a

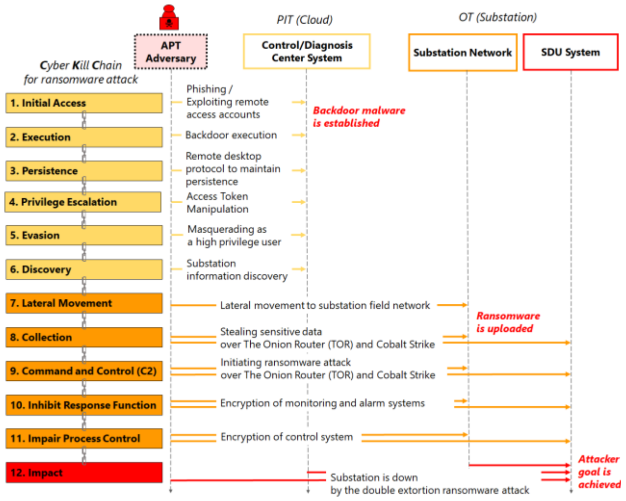

PIT (Cloud)
OT (Substation)
APT Adversary
Control/Diagnosis Center System
Substation Network
SDU System
Cyber Kill Chain for ransomware attack
1. Initial Access
2. Execution
3. Persistence
4. Privilege Escalation
5. Evasion
6. Discovery
7. Lateral Movement
8. Collection
9. Command and Control (C2)
10. Inhibit Response Function
11. Impair Process Control
12. Impact
Phishing / Exploiting remote access accounts
Backdoor malware is established
Backdoor execution
Remote desktop protocol to maintain persistence
Access Token Manipulation
Masquerading as a high privilege user
Substation information discovery
Lateral movement to substation field network
Ransomware is uploaded
Stealing sensitive data over The Onion Router (TOR) and Cobalt Strike
Initiating ransomware attack over The Onion Router (TOR) and Cobalt Strike
Encryption of monitoring and alarm systems
Encryption of control system
Attacker goal is achieved
Substation is down by the double extortion ransomware attack

Fig. 2. A cyber kill chain model for a substation ransomware attack.

valid remote desktop protocol (3. Persistence). The malicious cyber actor is manipulated an access token to get ownership of a malicious running process (4. Privilege Escalation). With the previous technique, the attacker is masqueraded himself as a high privilege user to avoid a detection system (5. Evasion). Afterward, the substation information is gathered from the PIT (6. Discovery). The APT actor can access the on-site SDU system with the field network authority, including all connected local devices (7. Lateral Movement). The APT collects field device data to learn the operation of the target substation. At this phase, malicious behaviors might be conducted over The Onion Router (TOR) and Cobalt Strike for anonymous communication (8. Collection). The ransomware file is loaded to the SDU system (9. Command and Control). By encrypting the response function, control process, and security alarm-related programs, the SDU controller is disabled to use (10. Inhibit Response Function and 11. Impair Process Control). Finally, the APT group demands the substation operator pay a ransom with an additional threat to disclose the collected substation data and system weakness information to the dark web (the double extortion ransomware attack). Finally, the substation can be shut down during the ransom negotiation period due to the encrypted software.

III. Proposed CNN-based Ransomware Attack Detection

This section describes a deep learning-based ransomware attack detection method. Fig. 3 illustrates the proposed deep learning algorithm designed by a CNN model using gray-scale image as an input to detect ransomware files. The design of ransomware detection method consists of three sequential processes: Data pre-processing, feature extraction, and classification. The proposed AI model can be implemented in the IDS in in Fig. 1 to prevent the malicious payload of malware files to the digital substation.

A. Data Preporcessing

Fig. 4 shows the data preprocessing of files to be viewed as 2-D image format. First, an executable ransomware or goodware file (i.e., binary file format) is converted to a vector form by reading unsigned 8-bit integers. A 2-D array is then created based on the size of the binary file [14].

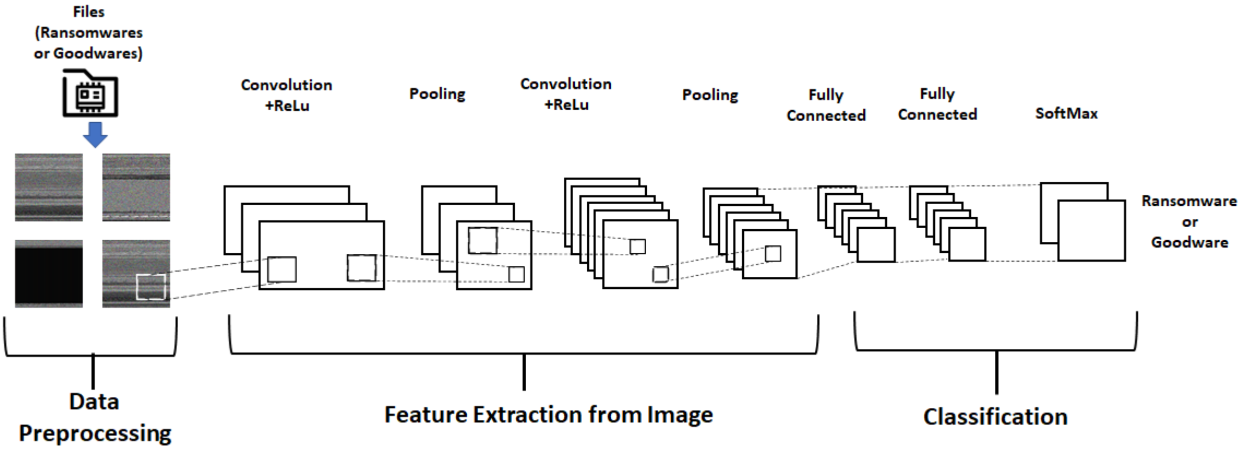


Fig. 3. Proposed CNN-based ransomware detection method.

Table I shows recommended Image widths according to the file size. After the matrix is formed, each value in the matrix is assigned in grayscale colors ranging from 0 to 255 (255: white, 0: black). After that, the 1-D grayscale array is converted to a 2-D grayscale matrix and, then it is finally converted to a 2-D grayscale image.

Unbalanced datasets are a common problem in computer vision and damage classification problems. The lack of images in each layer can lead to underfitting and overfitting, which has a big impact on CNN performance. Therefore, an data augmentation method is provided in the ransomware dataset to improve the performance of the classifier. Data Augmentation is a method used to enhance a dataset commonly used to train neural networks. In the enhancement phase we generate new data from classes with less population in the datasets. This process overcomes the limited impact on the data to avoid the unequal representation. To further improve the attack detection rates within limited number of ransomware samples, common preprocessing techniques such as rescaling and sample-wise standardization are also applied.

B. *Design of CNN Model for Ransomware Detection*

The CNN model architecture has a multi-layered structure consisting of two convolutional layers (CLs), two Max pooling layers (MPLs), two fully connected layers, and Softmax. CLs and PLs are used to extract multiple features from the preprocessed inputs. ReLu is chosen as the activation function as it does not change the size of the image and ReLU is good for increasing non-linearity from images which have high non-linearity. The noise impact of the features is reduced by the pooling layers. Two fully connected layer utilizes the output from the convolution process and predicts the class of the image based on the features extracted in previous stages. Each neuron is processed by a point element between small regions and weights related to the amount of information. Softmax is used in the layer of CNN which normalizes the CNN output between 1 (i.e., ransomware) and 0 (i.e., goodware). User-configurable hyperparameters including learning rate, number of hidden layers, number of hidden nodes, number of epochs, stack size, and type of activation function are optimally chose by trial-and-error effort.

TABLE I
IMAGE WIDTH ACCORDING TO VARIOUS FILE SIZE

File Size Range	Image Width
<10kB	32
10 kB – 30 kB	64
30 kB – 60 kB	128
100 kB – 200 kB	256
200 kB – 500 kB	384
500 kB – 1000 kB	512
>1000 kB	1024

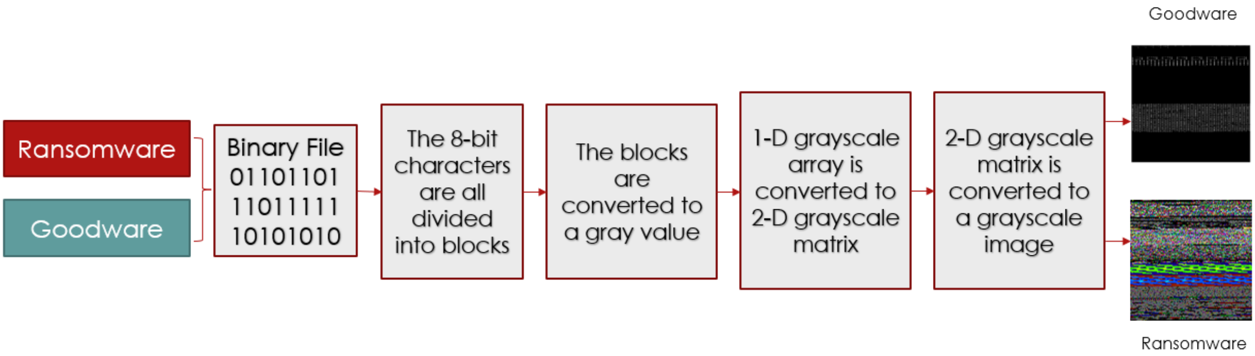
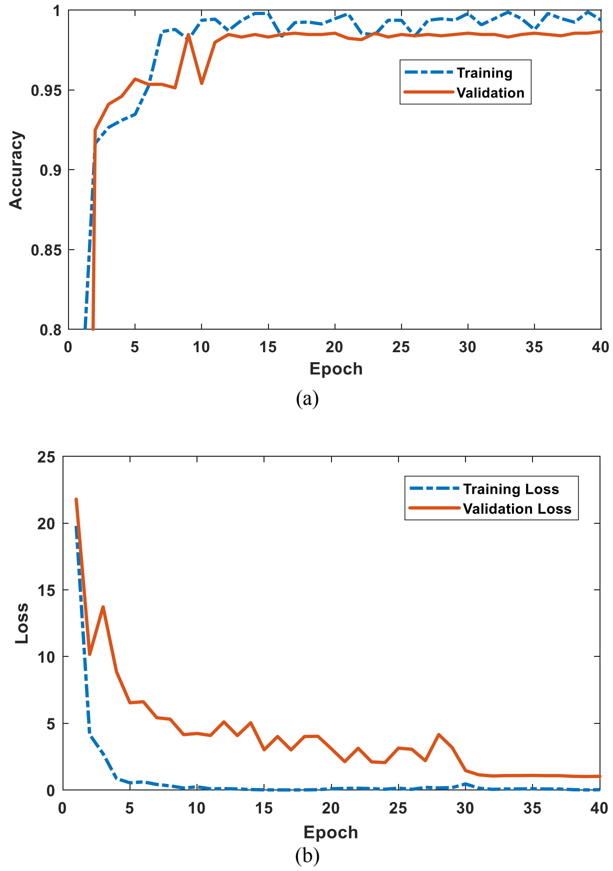

Goodware
Ransomware
Goodware
Binary File
01101101
11011111
10101010
The 8-bit characters are all divided into blocks
The blocks are converted to a gray value
1-D grayscale array is converted to 2-D grayscale matrix
2-D grayscale matrix is converted to a grayscale image
Ransomware

Fig. 4. Data preprocessing.

IV. Validation

The dataset consists of 672 goodware samples and 845 ransomware samples. The ransomwares files consist of five different families: Cerber, TeslaCrypt, Locky and Darkside. The goodware portable executable (PE) files are collected from windows platform and the from Portable Apps platform [15]. The datasets are split into two for training and testing purposes. The model is trained with 90% of the real ransomware file samples and augmented samples. The proposed CNN-based ransomware detection model is designed and trained in the COLLAB a cloud computing platform provided by Google. The experiment is run on Windows computer running i7 9750H, RTX 2060, 16GB RAM. The program is written in Python 3.9.7 with Kera's and PyTorch as backend.

Fig. 5(a) depicts the accuracy results of the training and experiments of the CNN-based detection model for 40 epochs. The accuracy values of training and validation converge to 99 % and 96.22%, respectively. The difference between these two accuracies is negligible. Fig. 5(b) shows the loss curves of training and validation converging to 0.0056 and 0.0388, respectively. This shows that the proposed model is unbiased for the training images, but also it provides high ransomware detection accuracy. These results show the proposed CNN-based detection method is accurate and suitable for the ransomware file detection.



(a)

(b)

Fig. 5. Training and Validation results of the CNN model: (a) accuracy and (b) loss.

TABLE I
The Comparison of Ransomware Detection Algorithms

Method	Feature Extraction	Datasets (ransomware/goodware)	Accuracy
Proposed	Images from raw files	672/845	96.22
Zhang et al. [16]	Opcodes from raw files using a disassembler	1787/100	91.43

Table II shows the comparison of the proposed CNN method and a RF method based on N-gram of opcodes which shows the best accuracy among ML classifiers among decision tree (DT), K-nearest neighbors algorithm (KNN), naive Bayes (NB), and gradient boosted decision trees (GBDT) [16]. Moreover, the opcode-based feature extraction technique requires a disassembler to get the opcode from the file, while the CNN-based method does not require the disassemble process by extracting features directly from the raw data. The comparison shows that the proposed CNN model using images provides better accuracy compared to the ML methods using op-codes.

V. Conclusion

This paper has explored potential attack surface of ransomware attacks in a digital substation and provided a CKC-based ransomware attack model. Moreover, this paper has investigated AI-based ransomware file detection methods. The CNN-based ransomware detection system can detect malware files with high accuracy without additional components such as disassembler. Future works include : 1) evaluating the proposed algorithm in the IDS in a real-time hardware-in-the loop (HIL) cybersecurity testbed for a smart substation and 2) improving the detection accuracy and reducing detection time.

References

[1] Alert (TA13-309A), CryptoLocker ransomware infections [Online]. Available: https://us-cert.cisa.gov/ncas/alerts/TA13-309A

[2] Kaspersky Lab, Kaspersky Security Bulletin, 2016.

[3] Alert (TA17-132A), Indicators associated with WannaCry ransomware, [Online]. Available: https://us-cert.cisa.gov/ncas/alerts/TA17-132A

[4] S. Larson and C. Singleton, "Ransomware in ICS environments," Dragos, Inc., White Paper, Dec. 2020.

[5] Novinson, M, Colonial pipeline hacked via inactive account without MFA. Accessed July 7, 2021, [Online]. Available: https://www.crn.com/news/security/colonial-pipeline-hacked-via-inactive-account-without-mfa

[6] [Online]. Available: https://www.virustotal.com/gui/home/upload

[7] B. Ahn, G. Bere, S. Ahmad, J. Choi, and T. Kim, "Blockchain-enabled security module for transforming conventional inverters toward firmware security-enhanced smart inverters," in *Proc. 2021 IEEE Energy Conversion Congress and Exposition*, Vancouver, Canada, Oct. 10-14, 2021, in press.

[8] A. Kharraz *et al.*, "Cutting the gordian knot: A look under the hood of ransomware attacks," in *Proc. Int. Conf. Detection of Intrusions and Malware and Vulnerability Assessment*, 2015.

[9] B. M. Khammas, "Ransomware detection using random forest technique," *ICT Express*, vol. 6, no. 4, pp. 325–331, Dec. 2020.

[10] G. Cusack, O. Michel, and E. Keller, "Machine learning-based detection of ransomware using SDN," in *Proc. the 2018 ACM International Workshop on Security in Software Defined Network & Network Funciton Virtualization*, Tempa, AZ, USA, Mar. 21, 2018, pp. 1–6.

[11] F. Maimo, *et al.*, "Intelligent and dynamic ransomware spread detection and mitigation in integrated clinical environments," *Sensors*, vol. 19.5, no. 1114, 2019.

[12] MITRE's ATT&CK for ICS, [Online]. Available: https://collaborate.mitre.org/attackics/index.php/Main_Page

[13] [Online]. Available: https://us-cert.cisa.gov/ncas/alerts/aa20-352a

[14] L. Nataraj, S. Karthikeyan, G. Jacob, and B. S. Manjunath, "Malware images: Visualization and automatic classification," in *Proc. 8th Invetenation Symposium on Visualization for Cyber Security*, Pittsburg, PA, USA, Jul. 20, 2011, pp.1–7.

[15] [Online]. Available: https://portableapps.com/apps.

[16] H. Zhang, et al., "Classification of ransomware families with machine learning based on N-gram of opcodes," *Future Gener. Comput. Syst*. vol. 90 pp. 211–221, 2019.