Image-Based Malware Classification Using Convolutional Neural Network

Hae-Jung Kim(✉)

Department of Cyber Security, Kyungil University,
Daegu 38428, Republic of Korea
hjkim325@kiu.kr

Abstract. In this paper, a malware analysis method that analyzes images learned by artificial intelligence deep learning to enable protection of big data by quickly detecting malware, including ransomware, is proposed. First, more than 2,400 datasets frequently used by malware are analyzed to learn and image data with a convolutional neural network. Data are then converted into an abstract image graph and parts of the graph extracted to find the group where malware exist. Through comparative analysis between the extracted subsets, the degree of similarity between these malware is analyzed experimentally. Fast extraction is achieved by using deep learning. Experimental results obtained indicate that use of artificial intelligence deep learning can enable fast and accurate malware detection by classifying malware through imaging.

Keywords: Malware · Artificial intelligence deep learning
Convolutional neural network (CNN) · Bit data security

1 Introduction

The Malicious software written for malicious malevolent purposes poses a serious threat to information security. With respect to information security for malware treatment, the malicious code must be correctly classified correctly. Moreover, ransomware, a type of malware that encrypts an infected PC and demands money, is expected to cause more cyber-attacks. In addition, the target of attacks by ransomware is changing from ordinary users to firms. Therefore, methods to protect big data from malware such as ransomware is becoming essential. And, a more accurate and easier method is required.

In this paper, an imaging method deep learning based malware inspection program that learns malware including ransomware and recognizes the pattern of malware using artificial intelligence deep learning is proposed. The proposed malware inspection method learns malware using artificial intelligence deep learning and use features that form similar images even if part of the malware changes. The performance of the proposed CNN model was the highest at 99.95%, indicating its viability as an accurate classification model.

J. J. Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 474,
https://doi.org/10.1007/978-981-10-7605-3_215

2 Background

2.1 Malware and Search Method

Malicious software includes all software that can have a malicious effect on computers, such as ransomware that make money. Ransomware are viewed as the worst type of malware because encrypted files cannot be restored even if this malware is eliminated through an antivirus program. Analysis and detection research on various malware is actively underway. However, malware such as ransomware have become progressively more intelligent and elaborate over time. Signature based detection is the method most frequently used in the antivirus industry to detect and counter current ransomware malware. The conventional malware detection method uses values hashed with an input binary file to search and detect identical hash values from a database. In this signature based malware detection system, problems occur when the hash value is different as even a change in a part of the malware can easily deceive the system.

Unlike the conventional signature based malware detection system, the method proposed in this paper images the malware to distinguish it even if a part of the malware changes. In addition, the search accuracy for malware and new mutations is increased by using artificial intelligence deep learning.

2.2 Related Research

In previous studies, the hash values of malware were used as signatures to search a database without classification or imaging malware. Consequently, previous methods could not accurately detect mutant malware from the database when hash values became completely different because part of the malware changed [1, 2].

In addition, machine learning based methods were used to learn features extracted from malware and features learned from new commands were searched in other studies [4]. Machine learning based classifiers that learn and verify extracted features using methods such as k-nn or XGBoost introduce fundamental preprocessing procedures and extracted features to machine learning methods. Functions of the classifiers were investigated by changing extracted features because the performances of classifiers substantially vary depending on the learning of features extracted from malware in these methods [5].

On the other hand, when malware are imaged, the images are used as features without selecting the features extracted from the malware. When too many features are used in malware detection systems that use machine learning based classifiers, the classification performance may degrade when unnecessary information is input in the classifier [6–8].

Recent deep learning methods that use artificial neural networks in the high resolution image field can successfully extract features of high-dimensional inputs that include unnecessary information. Therefore, deep convolutional neural network (CNN), which shows outstanding image classification performance in artificial intelligence, is used in the proposed system that learns malware and searches the imaged malware [9, 10].

3 Proposed Method

3.1 Malware Dataset

The Microsoft malware classification challenge dataset, which was used as the learning and verification data for the proposed artificial intelligence deep learning based malware detection system, was presented at the Kaggle machine learning challenge a machine learning based data analysis contest hosted by Microsoft in 2015 [3]. A total of 10,868 malware comprising nine different types and approximately 200 GB are shown in Table 1.

Table 1. Malware type and features in the data set

Glass index	Malware name	Description
1	RAMIT	Strong botnet function
2	Good Similar	Very well
3	KELIHOS v.3	p2p botnet using polymorphism Encrypted
4	VUNDO	Multi-component malware family: trojan, worm
5	SIMDA	Most complex malware, Multi-component malware family: botnet, trojan, backdoor, password-stealing
6	TRACUR	Trojan
7	KELIHOS v.1	Botnet
8	OBFUSCATOR. ACY	Combination of methods: Encryption, Compression, Anti-debugging, Anti-emulationtechniques
9	GATAK	Trojan

3.2 Structure of the Proposed Model

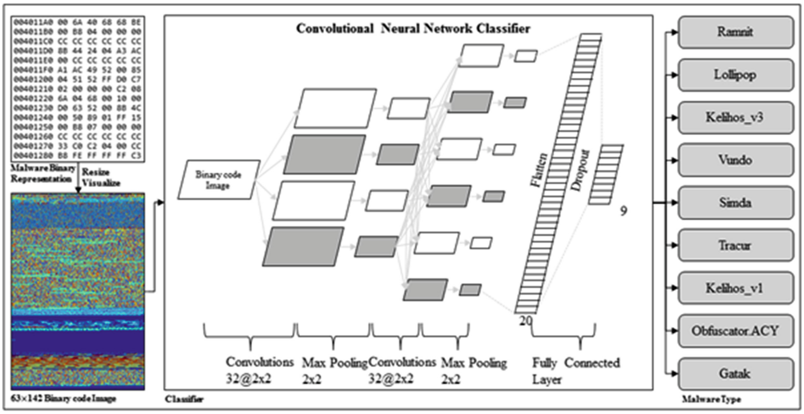
In studies that use classifiers based on machine learning to detect malware, only part of the features are used because features extracted from the malware are selected. Therefore, a malware detection system that uses artificial intelligence deep learning based deep CNN to automatically select the optimal features of high-dimensional images is proposed in this paper.

A CNN extracts optimal feature maps by performing convolution and pooling calculations from an image. Its operation is defined by the equation below.

$$V'^{1}_{xy} = \sum\nolimits_{a=0}^{m-1} \sum\nolimits_{b=0}^{m-1} w_{ab} y^{l-1}(x+a)(y+b) \tag{1}$$

Element V'_{xy} of vector V', which is the output from the first convolutional layer in Eq. (1), performs 3D convolution a layer y^{l-1}, which is the output vector of the previous layer, and filter w, which is an $m \times m$ vector. Max-pooling calculation, which selects one maximum value as the representative from a $k \times k$ region of the input $N \times N$ vector, is performed in the pooling layer, and a $\frac{N}{k} \times \frac{N}{k}$ vector is output.

Figure 1 shows the overall structure of the proposed method. Repetitive 2×2 convolution and 2×2 pooling calculations are performed on the malware of the imaged left binary file. Finally, one fully connected layer is used. Design using convolution and pooling is mainly employed in image classification, and the performance was proven to be the best in the contest. The type of malware is output through the proposed system when the imaged malware is input through learning.

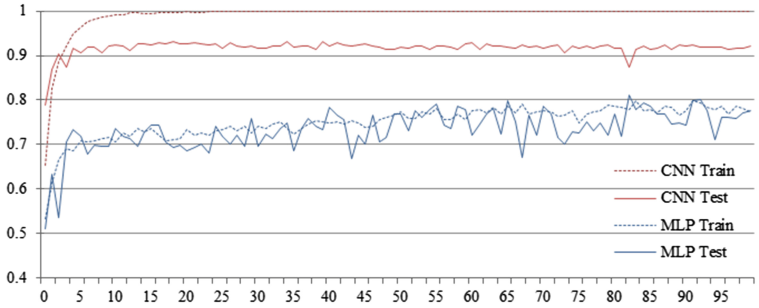


```
004011A0 00 6A 40 68 68 BE
004011B0 00 B8 04 00 00 00
004011C0 CC CC CC CC CC CC
004011D0 8B 44 24 04 A3 AC
004011E0 00 CC CC CC CC CC
004011F0 A1 AC 49 52 00 85
00401200 04 51 52 FF D0 C7
00401210 02 00 00 00 C2 08
00401220 6A 04 68 00 10 00
00401230 D0 63 52 00 8B 4C
00401240 00 50 89 01 FF 15
00401250 00 B8 07 00 00 00
00401260 CC CC CC CC CC CC
00401270 33 C0 C2 04 00 CC
00401280 8B FE FF FF FF C3
```

Fig. 1. Proposal malware inspection system and visual

4 Experimental Results

The proposed classifier was compared with a classifier that utilizes the ordinary artificial neural network method. The performance with the deep learning based CNN classifier was found to be superior. Internal parameters are differentiated and nodes are updated into values to minimize the loss function in the learning process and values of other nodes can be reflected to cause overfitting problems. Therefore, accuracy of learning and verification was shown via repetition to investigate overfitting of the proposed model in Fig. 1. The learning and verification accuracy of the conventional machine learning multilayer perceptron (MLP) method was unstable and varied in the range 70–80% as the repetitive learning proceeded. In contrast, the learning accuracy of the CNN method converged to virtually 100% after 15 repetitions and verification accuracy also showed stable tendency in the range 91–92%. Figure 2 shows the distribution of the data inside the proposed model using the t-SNE algorithm. When data are input into the CNN, the data are multiplied with the parameter of each node as they reach the deeper layer and the level of learning inside the model is achieved via grouping through visualization. This shows that quick classification by artificial neural network learning is possible to detect new evolving malware.



Epochs	10	20	30	40	50	60	70	80	90	100
CNN Train Accuracy	0.9863	0.9967	0.9988	0.9987	0.9989	0.9993	0.9991	0.9995	0.9996	**0.9995**
CNN TestAccuracy	0.9056	0.9306	0.9185	0.9139	0.9139	0.9139	0.9185	0.9213	0.9130	**0.9176**
MLP Train Accuracy	0.7129	0.7110	0.7422	0.7537	0.7610	0.7678	0.7681	0.7765	0.7831	0.7803
MLP Test Accuracy	0.6963	0.6944	0.6972	0.7417	0.7157	0.7861	0.7667	0.7491	0.7472	0.7722

Fig. 2. CNN, for MLP-based learning and accuracy by verifying repetitions

5 Conclusions

Conventional methods that detect malware in attached files or included in mail simply search the registry and compare changes to files in the database. With algorithms that use hash values, it is difficult to find accurate values because these algorithms are sensitive to changes in the hash value of the input. Therefore, it is difficult to distinguish real mail and they may be deceived, enabling malware such as ransomware when a massive amount of mail is received, even when an antivirus program is used. Artificial intelligence deep learning recognizes the pattern of malware and learns from data autonomously. In addition, data are imaged to accurately detect malware without being sensitive to small values. Moreover, this method can protect big data from malware. Substantial amounts of labeled malware data for deep learning research already exist and studies can be conducted using these data. However, the regulation on use of big data should become freer in our country. Malware are imaged and their size modified for learning and classification through deep learning based deep CNN in this paper. By using inputs modified to 1/10 size by imaging binary files, malware were distinguished with 91.7% accuracy. In addition, a comparative experiment was conducted with an MLP classifier and the learning and verification accuracy visualized on a graph in 2D space by repetition. Security can be strengthened by enhancing accuracy of malware search. By expanding malware search on the web using artificial intelligence in additional research, big data can be protected by applying protection for malware infection that can occur on social networking service (SNS) through smartphones.

References

1. Luo, X., Liao, Q.: Awareness education as the key to ransomware prevention. Inf. Syst. Secur. **16**(4), 195–202 (2007)
2. Vinod, P., Jaipur, R., Laxmi, V., Gaur, M.: Survey on malware detection methods. In: Proceedings of the 3rd Hackers' Workshop on Computer and Internet Security, pp. 74–79, March 2009
3. https://www.kaggle.com/c/malware-classification
4. Kumar, A., Sharma, N., Khanna, A., Gandhi, S.: Analysis of machine learning techniques used in malware classification in cloud computing environment. Int. J. Comput. Appl. **133**, 15–18 (2016)
5. Ahmadi, M., Ulyanov, D., Semenov, S., Trofimov, M., Giacinto, G.: Novel feature extraction, selection and fusion for effective malware family classification. In: Proceedings of the 6th ACM Conference on Data and Application Security and Privacy, pp. 183–194 (2016)
6. Nataraj, L., Karthikeyan, S., Jacob, G., Manjunath, B.S.: Malware images: visualization and automatic classification. In: Proceedings of the 8th International Symposium on Visualization for Cyber Security, p. 4 (2011)
7. Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. J. Mach. Learn. Res. **3**, 1157–1182 (2003)
8. Dy, J.G., Brodley, C.E.: Feature selection for unsupervised learning. J. Mach. Learn. Res. **5**, 845–889 (2004)
9. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
10. Sainath, T.N., Mohamed, A.R., Kingsbury, B., Ramabhadran, B.: Deep convolutional neural networks for LVCSR. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 8614–8618 (2013)