Mobile malware detection using deep neural network through opcode

Mohd Zaki Mas'ud[1,*], Noor Azleen Anuar[1], Nor Azman Mat Ariff[1], Nazrulazhar Bahaman[1] , Erman Hamid[1]

[1]Information Security, Digital Forensic and Computer Networking (INSFORNET),
Faculty of Information Technology and Communication,
Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia

[*]Corresponding e-mail: zaki.masud@utem.edu.my

Keywords: Mobile Malware Detection, Opcode, Deep Neural Network

ABSTRACT – Enhancement in mobile technology lead to rapid adoption of mobile devices among community at large and make mobile devices being used to collect and stored a large amount of sensitive information. This scenario has turn mobile devices as one of the malware attack's target. The expeditious growth of mobile malware has outweighed signature-based detection approach as it depends on a constant signature update and limited to known mobile malware. On the other hand, anomaly-based detection approach able to overcome this issue, yet this approach generates false alerts. One of the methods to improve the false alert drawback in anomaly-based detection is by introducing Deep Neural Network which able to learn on large number of features without having to go explicitly into optimizing the feature selection. This paper explores the mobile malware detection using Deep Neural Network approach through opcode analysis in classifying between benign and malicious mobile malware application.

1. INTRODUCTION

Advancement in mobile technology and connectivity has encourage community to used mobile devices daily for various activity. However, the existence of mobile devices had given attackers to exploit it by injecting malware into the devices. Currently mobile malware has the ability of being unrecognized in the application store [1], thus led users to be tricked into installing the infected programs. Once infected, malware started to infiltrate and replicate, hence causes further damage and loss of sensitive data. According to the 2018 Internet Security Trend Report issued by Symantec Corporation, [2] the discovery of latest mobile malware variations has significantly increased over the course of three years. In 2017, the discovery of new mobile malware had surged up to 54% compared to 2016. Even though in 2018, there is a slide decrease in the number of new mobile malware the impact of mobile malware attacks is still significantly high. This statistic implicates that a study in mobile malware detection is important to overcome mobile malware infection.

Mobile malware detection detects Mobile malware through the behaviors of the application and can be traced based on several features from the static or dynamic analysis. Dynamic analysis based on the traces during the activation of a mobile application whereas static analysis is based on the traces of mobile application binary. Dynamic analysis requires mobile application to be executed and the traces are captured live, these requires a huge processing power and storage to collect and stores the traces log. On the other hand, Static analysis can trace mobile application behavior by just analyzing the binary file, which is the *.apk* file. Static analysis approach requires lesser computational resources [3], making it the best approach to be adopted in mobile devices. One of the traces that can be observes in analyzing mobile application behaviors is opcode or bytecode of mobile application binary. As stated by [4] and [5], opcode or operation code act as a set of machine language instructions. These instructions are used in initiating certain operation in a computerized system [6], thus able to be used for detecting for any malicious operation.

Machine learning technique is one the common choice for researchers in developing or proposing a better and efficient malware detection. However, with the increasing number of data cause the traditional machine learning algorithm to come to a stagnant point. Nonetheless, an emergent machine learning technique called deep learning techniques can process huge amount of data efficiently and better than traditional machine learning technique. This research only discusses one of the Deep learning approaches which called as Deep Neural Network. Deep Neural Networks (DNN), which employ deep architectures in Neural Networks (NN), can represent functions with higher complexity if the numbers of layers and units in a single layer are increased [7].

The objective of this research paper is to explore the implementation of Deep neural network in mobile malware detection through the opcode analysis. The remainder of the paper will be structured as follows. Section 2 explains the research methodology used in this study. Section 3 presented the discussion on result and analysis of the proposed implementation. Ultimately, this paper is concluded in Section 4.

2. METHODOLOGY

This research paper implement a DNN algorithm for classifying between benign and malicious android application according to its opcode. An experiment was done by examining a total of 2,000 benign and malicious samples. 1000 malicious .apk were downloaded from the Android Malware Dataset shared by ArgusLab [8] and 1000 benign software was downloaded from the Google PlayStore.

The research methodology used in this research consist of three phases, known as the data acquisition phase, data preparation phase and implementation phase. The research methodology is illustrated in Figure 1.

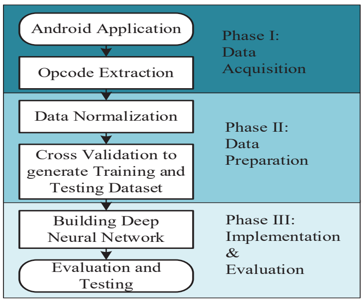
Android Application
Opcode Extraction
Phase I: Data Acquisition
Data Normalization
Cross Validation to generate Training and Testing Dataset
Phase II: Data Preparation
Building Deep Neural Network
Evaluation and Testing
Phase III: Implementation & Evaluation

Figure 1 Research Methodology

Phase I involved a set of malicious and benign .apk that are decompiled into Opcode occurrence. Opcode features are selected as it is able to display malicious traits even when there is encryption or obfuscation in the source codes. In Phase II, the opcode are normalized and divided to training and testing dataset. Finally, phase III is the implementation of DNN and the evaluation of DNN model .

The DNN used in this research have the following generic architecture: an input layer, which is provided with the opcode occurrence input, layer; 2 or more hidden layers, where a transformation is utilized to the output of the former layer and finally an output layer that classify between benign and malicious application. The architecture of the deep neural network implemented in this paper is depicted in Figure 2.

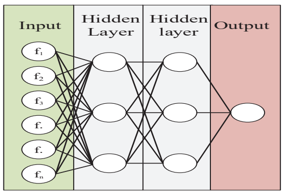

Figure 2 Deep Neural Network

3. RESULT AND DISCUSSION

This experiment is built using Keras utilizing TensorFlows backend and running on top Google Colab platform. The dataset of opcode occurrence is split into 60% training set and 40% validation set. The performance metric used to evaluate the performance of the classifier (malicious or benign) are Accuracy, Precision, Recall and F1-score. Table 1 describes the structure of the deep neural network. Relu and sigmoid function are used as activation functions.

Table 1 DNN parameters

Layer (Type)	Output Shape	Param #
dense_1 (Dense)	(None,110)	23980
dense_2 (Dense)	(None,110)	12210
dense_3 (Dense)	(None,1)	111
Total params : 36, 301 Trainable params : 36,301 Non-trainable params : 0		

The evaluation performance result of DNN in classifying benign and malicious mobile application is shown in Table 2.

Table 2 Performance result of DNN

Accuracy	Precision	Recall	F1-score
97.38 %	96.43%	98.18%	97.30%

Table 1. shows the ability of the proposed DNN to identify a mobile application as benign or malicious is 97.38%. The correctly classified apps within all application is 96.43%. The ratio of correctly identified benign samples is 98.18%. The F1-Score which points to how much the model determinate is 97.30%.

4. CONCLUSIONS

This research paper propose a mobile malware detection system using deep neural network through the opcode analysis, . The experimental result shows that deep neural network (DNN) can identify malicious from benign ones with 97.38% accuracy. In the future, this research will explore other Deep Learning method such as convolutional neural network (CNN) and Recurrent Neural Network (RNN)

ACKNOWLEDGEMENT

The authors are grateful to InforsNet Research Group of Universiti Teknikal Malaysia Melaka (UTeM) for the support and special acknowledgement to Ministry of Education Malaysia for providing financial support through the Fundamental Research Grant Scheme (FRGS/2018/FTMK-CACT/F00391).

REFERENCES

[1] Tenenboim-Chekina, L., O. Barad, A. Shabtai, D. Mimran, L. Rokach, B. Shapira, and Y. Elovici., Detecting application update attack on mobile devices through network feature, *IEEE Conference on Comp. Comm. Workshops*, pp. 91-92, 2013.

[2] Symantec Corporation, *2018* Internet Security Trend Report, [online] Available at: https://docs.broadcom.com/doc/istr-23-2018-executive-summary-en-aa [Accessed 7 Sep. 2019].

[3] V. P., A. Zemmari, and M. Conti, A machine learning based approach to detect malicious android apps using discriminant system calls *,Futur. Gener. Comput. Syst.*, vol. 94, pp. 333–350, 2019.

[4] P. Faruki et al., Android sec.: A Survey of issues, malware penetration, and defenses, *IEEE Commun. Surv. Tut.*, vol. 17, no. 2, pp. 998–1022, 2015.

[5] S. Rezaei, A. Afraz, F. Rezaei, and M. R. Shamani, Malware detection using opcodes statistical features, *8th Int. Symp. on Telec. (IST)*, 2016, pp. 151–155.

[6] P. Feng, J. Ma, C. Sun, X. Xu, and Y. Ma, A novel dynamic android malware detection system with ensemble learning, *IEEE Access*, vol. 6, pp. 30996–31011, 2018.

[7] Liu, Weibo, Zidong Wang, Xiaohui Liu, Nianyin Zeng, Yurong Liu, and Fuad E. Alsaadi. A survey of deep neural network architectures and their applications. *Neurocomputing* 234 (2017): 11-26.

[8] F. Wei, Y. Li, S. Roy, X. Ou, and W. Zhou, Deep ground truth analysis of current android malware, *Lect. Notes in Comp. Sci.*, vol. 10327, 2017, pp. 252–276.