Formulating ensemble mobile malware detection through n-gram system call sequence features

Nor Azman Mat Ariff, Mohd Zaki Mas'ud*, Amizah Aida Ahmad, Nazrulazhar Bahaman, Erman Hamid

[1]Fakulti Teknologi Maklumat dan Komunikasi, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia

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

Keywords: Mobile malware detection; ensemble method; N-gram

ABSTRACT – The extensive usage of mobile devices among the community has turned the mobile technology as one of the driven force in the increase of malware attacks. The rapid evolution of mobile malware has overcome the malware signature detection approach. This approach requires a constant signature update and only able to detect a known mobile malware. Anomaly detection approach can overcome this issue but using a single classifier can degrade the classification accuracy. Based on this reason, this research formulate an ensemble of different n-gram system call sequence features to improve the accuracy of mobile malware detection. However, the determination of the best number of sequence (n) to be used as the features have become an issue. In order to resolve this drawbacks, this research introduces an approach that applied different n-gram sequence call feature to construct n number of classifier models. The probability output of each classifier is then combined to produce a better prediction output that determines the mobile application is benign or malicious. The combination of this multiple-classifier produces a better predictive performance compared to a single classifier.

1. INTRODUCTION

The proliferation of mobile devices in recent years has been increasing rapidly. The functionality of a single mobile phone has become similar to the desktop computer. In the end of 2016, the International Telecommunication Union (ITU) [1] forecasting that there are 7.5 billion mobile users and 3.8 billion mobile‐broadband subscribers worldwide. In the meantime, the Gartner Inc. [2] reported, 84.1 % of the smartphone market has been monopolized by the Android based mobile device in the first quarter of 2017. Regardless the popularity and the technological advancement of the Android-based mobile possessed in the market, nowadays it has become the main target of malware author. The number of mobile malware targeting android based platform has increased tremendously over the year. In 2012 alone, Kaspersky Security Bulletin [3] has reported that 98.96% of newly found mobile malware is targeting Android.

Antivirus for mobile detects malware based on the known malware signature. However, with a more advanced malware introduced, the signature is kept on changing from one variant to another variant, thus making it difficult to detect a new malware unless the signature database is updated. Updating the signature database led to a larger signature database, thus acquiring more storage and higher memory consumption to compute. In contrary, the classification approach in anomaly-based mobile malware detection have the ability to detect previously unknown malware [4]. Nevertheless, [5] have shown that single classifier has their own domain of competence, therefore is not an optimal approach to solve all problems. This limitation leads to the ensemble methods which exploit the strengths of individual classifier models by performing information fusion of classification decisions. Ensemble methods train multiple classifiers to solve the same problem. Ensemble methods construct a set of classifiers and combine them.

One of the proposed solutions to enhance mobile malware detection is to use n-gram system call sequence as a feature in classifying benign and malicious mobile applications [6]. Tracing each element of mobile malware behaviour on the captured system call log has discovered that each element has its own set of system call processes and a sequence of system call invoke that can represent a malicious mobile application process. The sequential system call is known as n-gram analysis, where n is the number of the sequence length or the number of co-occurring sets of system call invoked in the process. Crowdroid [7], and Dini et. al [8] are among the works using system call as the features in classifying benign and malicious mobile application, yet the approach only takes each single occurrence of the system call or 1-gram as the feature. Based on this, the research is considering several n value to be the classification features to improve the accuracy.

2. METHODOLOGY

Each n-gram system call sequence is applied to a respective Support Vector Machine (SVM) classifier and the probability output from each SVM classification decisions are combined using bagging ensemble method. Bagging exploits the independence between the individual classifiers, since the error can be reduced dramatically by combining individual classifiers. Each classifier is trained on each subset and combines them using several combination methods. The classification accuracy can significantly improve if the error of the single classifier is not strongly correlated.

For this research 125 malicious and 101 mobile applications are used. The sample is randomly sub-sampling into train and test sets. Each set consists of 100 instances with 50 malware and 50 non-malware

instances. The experiment is repeated 10 times and the accuracy is calculated by averaging the results from all 10 runs. The experiments contain two phases. In phase 1, the single classifier models for 1-gram to 5-gram are generated. The representation for 1-gram to 5-gram feature vectors using bag-of-words (BOW) model, information gain (IG), symmetrical uncertainty (SU), and chi-square (CS). Based on the results of the phase 1, the experiment is extended to formulate an ensemble method that combine five BOW n-gram models. Three combination methods, product rule, mean rule and majority voting are used to evaluate the performance of the newly formulated ensemble method. Figure 1 shows the experiment process.

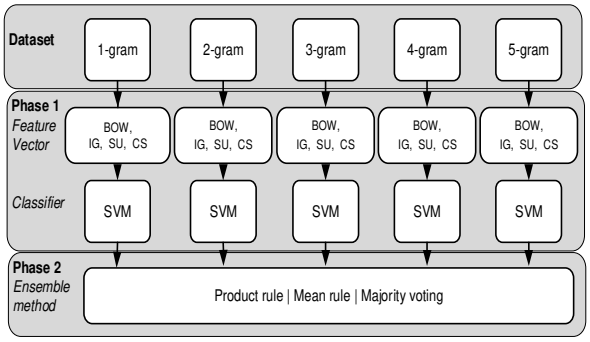

Figure 1 Experiment process flow

3. RESULT AND DISCUSSIONS

Table 1 shows the average classification accuracy (%) for individual BOW classifier models with IG, SU and CS feature selection applied

Table 1 The average classification accuracy of the n-gram features (n=1 to n=5)

n-gram	BOW	IG	SU	CS
1	85.8	85.4	85.3	85.3
2	**92.4**	92.0	92.2	92.3
3	90.3	89.6	90.2	90.2
4	90.4	90.5	90.5	90.4
5	89.9	89.9	89.8	89.8

The table shows the best performance is obtained at BOW 2-gram model with the average accuracy 92.4%. Feature selections give average results and cannot outperform BOW. In this case, we believe that BOW itself has sufficiently information to classify the malware and non-malware.

Table 2 The average classification accuracy using ensemble methods

BOW	Product Rule	Mean Rule	Majority Voting
92.4	**98.6**	98.5	96.9

Table 2 shows the average classification accuracy (%) for ensemble method. Best result from first stage, which is 92.4 % for BOW 2-gram model is compared with the proposed ensemble method. The table clearly shows that proposed ensemble method significantly outperform the best individual classifier. The best performance is achieved when the product rule combination method is applied. However, majority voting give a lower accuracy result compared to product rule and mean rule, but better than BOW 2-gram model.

4. CONCLUSIONS

This research, proposed an anomaly approach mobile malware detection approach through an ensemble n-gram system call sequence. The classification result shows a bagging ensemble classifier using the the product rule combination method gives the highest accuracy in classifying between malicious and benign mobile application. Thus proof that a combination of multiple classifier is better than considering only single classifier.

ACKNOWLEDGEMENT

Authors are grateful to Universiti Teknikal Malaysia Melaka and Ministry of Higher Education of Malaysia for the financial support through Fundamental Research Grant Scheme FRGS/1/2015/ICT04/FTMK/02/F00290 and UTeM's Short Grant Scheme PJP/2018/FTMK(4B)/S01631.

REFERENCES

[1] International Telecommunications Union (ITU), 2016. ICT Facts Figures 2016. [ONLINE] Available at http://www.itu.int [Accessed 2 February 2017]

[2] Forni A. A. & Meulen R. V. D., 2017. Gartner Says Worldwide Sales of Smartphones Grew 9 Percent in First Quarter of 2017. [ONLINE] Available at http://www.gartner.com/newsroom/id/3725117. [Accessed 30 May 2017]

[3] Denis Maslennikov and Yury Namestnikov, Kaspersky Security Bulletin 2012. The overall statistics for 2012, (Securelist), [ONLINE] http://www.securelist.com/en/analysis/204792255/Kaspersky_Security_Bulletin_2012_The_overall_statistics_for_2012#1. [Accessed 25 July 2013]

[4] Moon, S. S. & Kyeong, J. J., 2006. Alert Correlation Analysis in Intrusion Detection. Proceedings of the 2nd International Conference Advanced Data Mining and Applications (ADMA 2006), pp. 1049–1056.

[5] Abdullah, Azizi, Remco C. Veltkamp, and Marco A. Wiering. Ensembles of novel visual keywords descriptors for image categorization. In Control Automation Robotics & Vision (ICARCV), 2010 11th International Conference on, pp. 1206-1211. IEEE, 2010.

[6] Mas' ud MZ, Sahib S, Abdollah MF, Selamat SR, Yusof R, Ahmad R. Profiling mobile malware behaviour through hybrid malware analysis approach. Information Assurance and Security (IAS), 2013 9th International Conference on 2013 Dec 4 (pp. 78-84). IEEE.

[7] Burguera, I., Zurutuza, U., & Nadjm-Tehrani, S., 2011. Crowdroid: behavior-based malware detection system for android. In Proceedings of the 1st ACM workshop on Security and privacy in smartphones and mobile devices (pp. 15-26). ACM.

[8] Dini, G., Martinelli, F., Saracino, A., & Sgandurra, D., 2012. Madam: a multi-level anomaly detector for android malware. In Computer Network Security (pp. 240-253). Springer Berlin Heidelberg.