I.J. Modern Education and Computer Science, **2022, 5, 24-33**
Published Online on October 8, 2022 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijmecs.2022.05.03

MiMaLo: Advanced Normalization Method for Mobile Malware Detection

Sriyanto
Institut Informatika dan Bisnis Darmajaya, Indonesia
E-mail: sriyanto@darmajaya.ac.id

Sahib B. Sahrin
University of Technical Malaysia Melaka, Malaysia
E-mail: shahrinsahib@utem.edu.my

Abdullah Mohd. Faizal
University of Technical Malaysia Melaka, Malaysia
E-mail: faizalabdollah@utem.edu.my

Nanna Suryana
Multimedia University, Malaysia
E-mail: nsuryana.herman@mmu.edu.my

Adang Suhendra
Universitas Gunadarma, Indonesia
E-mail: snsebatik@gmail.com

Received: 02 March 2022; Accepted: 13 May 2022; Published: 08 October 2022

Abstract: A range of research procedures have been executed to overcome malware attacks. This research used a malware behavior observe approach on device calls on mobile devices operating gadget kernel. An application used to be mounted on mobile gadget to gather facts and processed them to get dataset. This research used data mining classification approach method and validates it using ten fold cross validation. MiMaLo is a method to normalize a dataset the usage of the min-max aggregate and logarithm function. The application of the MiMaLo method aims to increase the accuracy value. Derived from the experiments, the classifiers overall performance level used to be extensively increasing. The application of the MiMaLo method using the neural network algorithm produces an accuracy of 93.54% with AUC of 0.982.

Index Terms: Malware Attack; Mobile Malware Detection; Normalization Methods; MiMaLo

1. Introduction

In recent decades, the development of mobile device technology particularly mobile phones is immensely rapid. Mobile phones that initially rely solely on the texting and speak features, nowadays it is already broaden and become a trend in the community. Even some features of smartphones have become a lifestyle for the community. Along with it, smartphone producers cut the prices, make it affordable, thus the explosion of smartphone competes human population.

The Android operating system is one of the operating systems most widely used today. This Linux-based operating system provides an open platform for developers to create their own applications to be used by a variety of mobile devices. The applications available in the market Google Android Play already reached approximately 2,673,850 apps early in 2017. According to the International Data Corporation IDC, Android OS dominates 85% of the total market share in Q1 2017.

Following this supremacy of operating system, the hackers increasingly provoke malicious activity through the spread of various types of malware. Malware is software that is explicitly designed to perform malicious activity or other devices, such as trojan destroyer, virus, spyware, and more malicious activity resulting from malware will adversely affect the personal information that is stolen when the victim, the system is destroyed, and the activity is tapped. This is a digital evidence of a crime committed by a hacker by utilizing malware as a medium Google already makes cloud-based security system that aims to detect harmful applications in Play Store, however, number of

applications containing malware are still on Play Store and numerous Android users download them.

Number of methods have been proposed to detect malware attack on mobile [1] proposes a dynamic malware detection approach based on a call tracking system. Although it provides a simpler feature encoding, Android-specific toolkit, and extensive empirical evaluations, it also comes with some encoding of the application's fingerprint behavior into features for subsequent classification. This is based on extensive empirical evaluations on a set of more than 12,000 Android apps and then analyzes how the quality of malware classifiers is affected across multiple dimensions, including the choice of coding system calls to features, relative size of the benign and malicious data sets used in experiments, classification, and the size and type of inputs that drive dynamic analysis.

Comput, et al [2], proposes a new approach to detecting malware by means of hybrids and generics, especially for mobile malware on Android devices. The way it done is by performing an analysis of the execution data of a series of malware instances and non-malicious applications to generate individual system call patterns and sequential system calls with different call depths associated with file and network access, and so on. By comparing the patterns reflected by the above individuals and calling sequential systems from malware and applications that are benign to each other, then built a set of malicious patterns and a set of normal patterns used for malware detection and benign application assessment. When there is a need to detect unknown applications, it uses dynamic methods to collect system call time data in both individual calls and sequential system calls at different depths. Then extract the target pattern, for instance, the frequency of the sequential system calls with different call depths of the unknown app from the system call data of the process time. By comparing between a set of harmful patterns and a set normal pattern, it can assess unknown apps good or bad.

To get high accuracy in the data mining classification process, it is necessary to process data pre-processing. One of the processes in data pre-processing is data normalization. However, not all data or algorithms can be implemented on all types of data sets, as is the case of malware detection. The existing data set has a high range value so it is necessary to design a special method to overcome it. Therefore, the MiMaLo method is proposed, namely a method that uses a minimum-maximum approach and a logarithm based on log 10. By using the MiMaLo method, it is expected that the existing data set can become more normal so that when the classification process is carried out it has a high accuracy value.

2. Literature Review

2.1. Malware Analysis Method:

1. Dynamic Analysis

In this method, application will be activated within a secure environment both on a physical machine that has been provided as a laboratory or in a virtual machine [3,4]. We can note any activity performed by malware when successfully infects a computer. Stages in the analysis of this dynamic will check the computer with the overall such a process running on your computer, change the registry, internet communications and other awkward events that allow occurs when a computer has infected by malware. The researchers who use the dynamic analysis approach are [5, 6].

2. Static analysis

In this method, malware will not be enabled by default except to trace and examined and analysed against the source code written in the malware program by performing surgical stages against the malware program, so that the information obtained is extremely complete and can give very detailed description about the mechanism of action of such malware as a whole. In using the method of static analysis of malware it is claimed capable of understanding the language of the machine, especially the architecture of a program as it will be very helpful in analysing the composition of a malware program code related to gather information from behaviour evoked by the malware [7, 8].

3. Hybrid Analysis

Hybrid analysis method is a combination of dynamic and static methods that is applying dynamic ways to collect system call data during application execution by modifying cellular OS and using static methods to analyze data to detect good and bad applications by processing data collected on computing servers [2, 9].

2.2. Mobile Malware Detection

The following information is the summary of the benefits and limitations of each mobile malware detection element that can be used to describe the requirements required in the mobile malware detection system framework.

1. Analysis approach categories:

a. Static Analysis approach categories have the advantage of quickly detecting malware and being able to prevent malicious applications from being installed but having weaknesses can be easily avoided through obfuscation or encryption techniques [10].

b. Dynamic Analyst approach category has advantages because with obfuscation and encryption techniques it will not affect detection performance but has weaknesses in terms of generating captured data generation, requires storage space, requires high computational power to process data, time needed to observe the activity appearance dangerous cannot be clearly defined [11].

2.Technical categories used:

a. Signature-based techniques have the advantage of being able to detect attacks that are known accurately and use less computing resources. The disadvantages are less effective for unknown or new malware [8, 10].
b. Anomal-based techniques have the advantage of being very effective in dealing with unknown but flawed malware requiring more computing resources and generate a number of false warnings [6].
c. Specifications-based techniques have the advantage of being very effective in dealing with unknown malware but having weaknesses in the event of a decrease in detailed specifications for SPB takes time and there is a possibility of detection will produce false negative [12].

3. Audit data sources used:

a. Application package advantages Less complex to analyze but have weaknesses can be avoided through obfuscation or encryption techniques and does not reveal the true behavior of malware [8, 10].
b. Sensitive data flow gains are less complex to analyze and have weaknesses because not all malicious applications process sensitive information [13].
c. Kernel level system calls have the advantage of being able to reveal the dynamic behavior of malware but require complex analysis and generate large amounts of data [12, 14].
d. Network Traffic has the advantage of being able to reveal the dynamic behavior of malware and have disadvantages due to complicated processes and generate large amounts of data [15].
e. Hardware has the advantage of being able to reveal the dynamic behavior of malware and has disadvantages because the process is complicated and generates large amounts of data but has abnormalities that can be affected by heavy user use [16].

Based on the analysis of benefits and restrictions on each element, mobile malware detection system framework must have the ability to:

a. Cellular Malware Analysis that can overcome all techniques of obfuscation and encryption and is able to process small amounts of input data captured with less computational complexity.
b. Mobile Malware Device Detection Techniques that can detect accurately and unknown known mobile malware with high detection accuracy while providing false warnings.
c. Audit data sources that can be used to produce high classification accuracy with only a small number of logs generated when dealing with obfuscation and encryption techniques.

2.3. Normalization

Variable transformations are applied to all values in a variable. In other words, for each object, the transformation is performed for the value of the variable for that object. For example, if only the magnitude of the variable is important, then the value of the variable can be transformed by setting the absolute value. There are two forms of variable transformation, namely:

1. Transformation of simple functions

In this transformation, simple mathematical functions can be applied to each individual value. If x is a variable, then the example of this transformation is xk, log x, ex, sin x, 1 / x, x0.5 and | x |. Transformation needs to be done carefully because it can change the nature of the data. For example, transformation 1 / x will reduce the value of the value equal to 1 or greater, but increase the value in the range 0 to 1. As an illustration, {1, 2, 3} is transformed to {1, ½, 1 / 3} while {1, ½, 1/3} is transformed to {1, 2, 3}. Thus, for all sets of values, the results of transformation 1 / x reverse the order of values.

2. Normalization or Standardization

Another form of variable transformation is standardization or normalization of variables. The goal is to create a set of values that have certain properties. One example is the standardization of internal variables statistics. If x is the mean or the mean of the attribute value and sx is the standard deviation, then the transformation x '= (x - x) / s x creates a new variable that has an average of 0 and a standard deviation of 1. There are several methods / techniques applied for data normalization, including:

1) Min-max Normalization

Min-max normalization maps a value v from attribute A to 'v in the range based on the (1)

$$v^l = \frac{v - \min A}{\max A - \min A} (new\_\max A - \ new\_\min A) + new\_\min A \quad (1)$$

2) Z-Score Normalization

Also called zero-mean normalization, where the value of an attribute A is normalized based on the average value and standard deviation of attribute A. The A value of v from attribute A is normalized to 'v with (2)

$$v^i = \frac{v - \bar{A}}{\sigma A} \quad (2)$$

3) Normalization by Decimal Scaling

Normalization is obtained by shifting the decimal point from the value of an attribute A. The number of decimal points shifted depends on the maximum absolute value of attribute A.

3. Proposed Method

Based on statistical analysis, the dataset used is mostly abnormal and has a wide range, for this reason it is necessary to normalize the process. One of the methods is Min-Max, the normalization method by performing linear transformations of the original data [17]. The formula for the min-max normalization method is (3)

$$newdata = (data - min)\ X\ \frac{(newmax - newmin)}{(max - \min\ )} + newmin \quad (3)$$

Where :

newdata = data resulting from normalization
min = minimum value of data per column
max = maximum value of data per column

In basic mathematics, the logarithmic number is the opposite of exponential. By using this logarithma, the value of the number can go down depending on what base logarithm is used. This principle is what researchers take to reduce the range of numbers in very high datasets [18]. In this study, the working principle of logarithms is combined with the min-max method on data normalization to get optimal performance.

This paper proposed a new method, namely MiMaLo, to improve the performance of the resulting higher model using data normalization approach. Based on the results of the analysis of the datasets that have been created, the data used is all in an abnormal or asymmetrical state. From the results of data processing using statistical techniques found data in the form of Kurtosis curves and has a very high range. If the data is used to create a data mining classification model, it will produce models that have poor performance. To improve the performance measured in the form of accuracy, recall, precision and Area Under Curve (AUC), the dataset must be normalized. Given the very high range, an approach with the Log function is used. Existing data will be processed with the log function then the results will be normalized using the Min-Max technique. The process series of the method with a logarithm and normalization function approach with Min-Max is so that this method is named MiMaLo. From the MiMaLo process it produces data that has a far lower range and approaches the normal distribution.

There are several feature selection methods available today such as Chi-Square, Information Gain, Relief, PCA and others, but from the results of preliminary research shows that some of these methods have not produced high-performance models for malware detection cases using this dataset.

3.1 Methods Improvement

Based on the analysis that has been carried out on the dataset, it is found that all existing data is abnormal or not symmetrical. Besides that, it was also found that the range of values between one attribute or feature and another was extremely high. Statistically, it can be said that the data has not been in an ideal state because it was not normally distributed. To overcome this, a new algorithm called MiMaLo algorithm is proposed. This algorithm is used to make the dataset to be more normal and reduce the data range. The algorithm is shown in Fig 1.

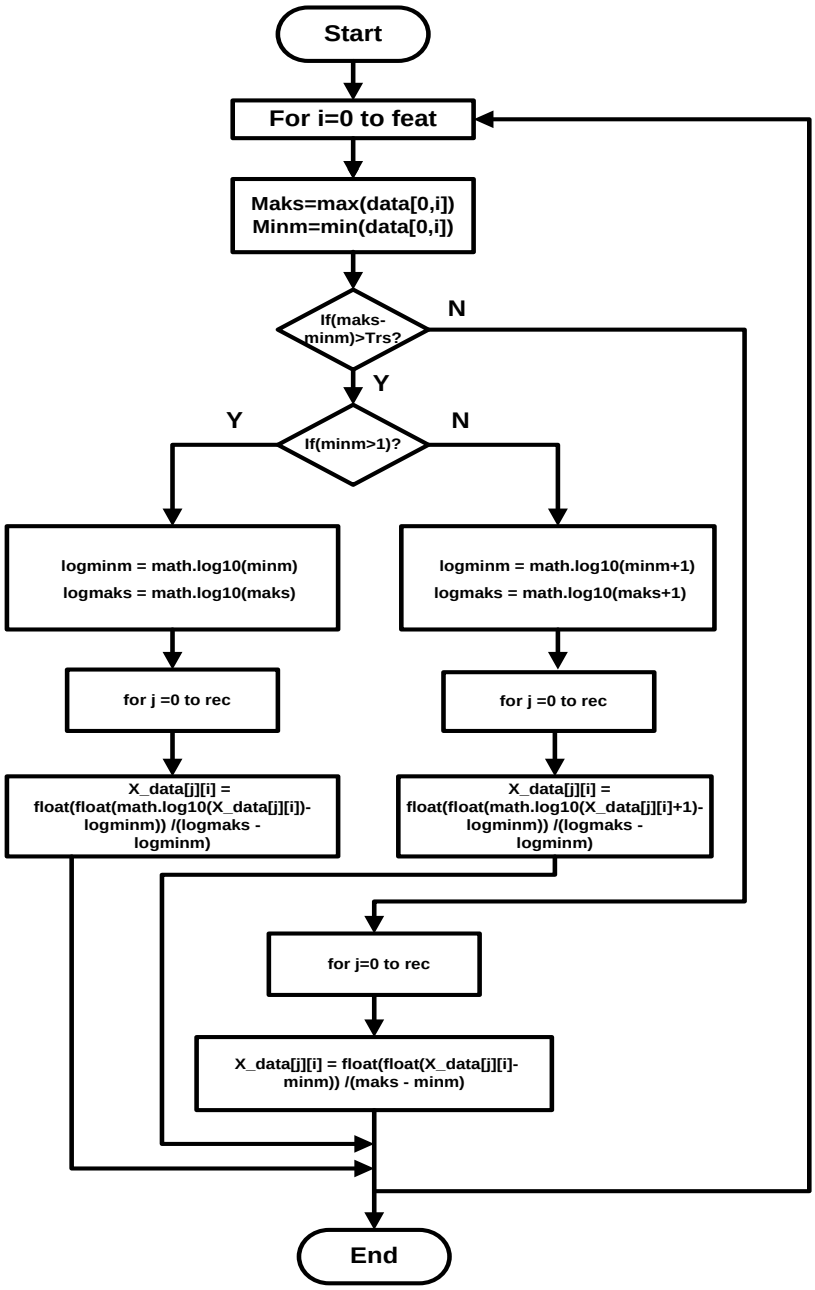

Fig.1. MiMaLo Algorithm

```
For i=0 to feat  logminm = math.log 10(minm)   logmaks = math.log10(maks)
for j =0 to rec YMaks=max(data[0,i]) Minm=min(data[0,i])YIf(maks-minm)>Trs
If(minm>1)
   X_data[j][i] = float(float(math.log10(X_data[j][i])-logminm)) /(logmaks - logminm)
   logminm = math.log10(minm+1) logmaks = math.log10(maks+1)
for  j  =0  to  rec      X_data[j][i]  =  float(float(math.log10(X_data[j][i]+1)-logminm))
/(logmaks - logminm)
for j=0 to rec  X_data[j][i] = float(float(X_data[j][i]-minm)) /(maks - minm)YNN
Start
End
```

Based on the above algorithm, it is illustrated that first the minimum-maximum value will be calculated and then the logarithm based on log 10 which will produce a more normal data set.

In this proposed method, the existing datasets will be processed using the MiMaLo algorithm to obtain more normal data. This research consist of five main phases: Data Preparation, Data Mining Modeling, Model Validation & Evaluation, Model Comparison and Results Analysis.

1. Data Preparation

1) Data Collection

In this phase the researchers conducted data collection to be used in research. The data will be taken in the form of log data in the android system. For this data retrieval process required a system that is specially configured to ensure that malware from the Internet network is free to infect mobile devices without being exposed to filters by the anti-virus system or firewall from the existing network. Equipment used in data retrieval is internet connection directly (without passing firewall), mobile devices (smartphones) and access point used to connect internet network. On smartphones installed a program used to record the activity of the system during the active application being used.

2) Data Preprocessing

a. Data Cleaning and Integration

Data cleaning is a process of removing data duplication, checking inconsistent data, and correcting data errors, such as input errors. In general, data obtained from either a company's database or experimental results, has an imperfect content such as missing data, invalid data or also just a typo. In addition, there are also data attributes that are not relevant to the data mining hypothesis. Data cleaning will also affect the results of information from data mining techniques because the data handled will decrease the number and complexity.

While Data Integration is the process of adding existing data with data or other information that is relevant or can be called is also a merging of data from various databases into a new database required. Cleaning and integration stages assume that data integrators must remove noise from initial data in parallel by integrating multiple data sets.

b. Data selection and Data Transformation.

Selection of relevant data and can be analyzed from operational data. The result data is stored in a separate database. While the data transformation stage is a necessary process for data to be used have the same format. The process of transforming the data into a certain format to make the data appropriates for the process of data mining. For example some standard methods such as association analysis and clustering can only accept categorical data input.

2. Data Mining Modeling

This stage is the application of data mining techniques for model building which will be carried out in the process of validation, evaluation and comparison. This research will use classification techniques using basic algorithms namely Decision Tree (DT), k-Nearest Neighbor (k-NN), Naïve Bayes (NB), Logistic Regression (LR), Neural Network (NN) and Support Vectors Machine (SVM). All proposed methods and existing methods will be applied to the algorithm to evaluate its performance.

3. Model Validation

The validation technique used is Ten Fold Cross Validation, where the existing dataset will be divided into two, namely as training data and testing data. Training data is used for learning algorithms to produce models, while testing data is used to test the resulting model. This division process is carried out ten times with different data composition. With validation techniques, it is not necessary to manually distribute training and testing data.

4. Model Evaluation

The evaluation stage is needed to determine the performance of the model that has been generated by the classifier. There are many ways that can be used to measure one of them using Confusion Matrix. Confusion matrix is a method usually used to perform accurate calculations on the concept of data mining. This method performs calculations by producing 4 outputs, namely: recall, precision, accuracy and error rate. The evaluation of a classification model is based on testing to estimate the right and wrong objects. To perform an analysis of the calculations that have been done can use ROC Curve, which is one way to analyze the model classifier that has been made. The use of ROC curves is to specify the model parameters to be desired in accordance with the characteristics of the classifier model. Classification methods can be evaluated based on criteria such as level of accuracy, speed, reliability, scalability and more.

5. Model Comparison

In this study also used the method of comparison test of parametric test (t-test) to compare the accuracy of classification algorithm. The accuracy value obtained is compared using the t-test to ascertain whether there are significant differences in algorithm accuracy. If the difference between two average accuracies are insignificant, it can be said that algorithmic accuracy can not be distinguished and if the difference is significant, then one algorithm has a poor accuracy compared to other algorithms. The smaller the value of the test results then it can be said that the algorithm is getting better. In this stage a comparison was made using the t-test between the basic datasets with the MiMaLo.

6. Result Analysis

At this stage an analysis of the results of several previous stages was carried out, namely the data preparation stage, data mining modeling, validation model, model evaluation up to the model comparison stage. At this stage, analysis of datasets, performance analysis of six classifiers using basic datasets and MiMaLo algorithm will be carried out.

4. Result And Discussion

Algorithm MiMaLo was proposed to improve the performance of basic method in detecting the malware by using dataset system call. This chapter contains the results of experiments include RO1, RO2, RO3 and RO4. The results include: the performance of six (6) basic classifiers based on syscall dataset (1 gram), the performance of six dataset-based classifiers resulting from the proposed method, the performance of the proposed feature selection method and performance comparison with other feature selection methods

4.1. Performance of Proposed Methods

Based on the experiments that have been carried out using a system calls dataset and continued with the application of the proposed method to improve the performance of the MiMaLo classifiers, the results are presented in table 1.

Table 1. Results of MiMaLo

#	CLASSIFIERS	ACCURACY	AUC	PRECISION	RECALL
1	DT	79.79	0.789	75.42	89.58
2	kNN	90	0.856	87.73	93.33
3	LR	90.21	0.961	84.2	99.58
4	NB	75.21	0.866	68.32	88.75
5	NN	93.54	0.982	90.51	97.5
6	SVM	88.54	0.96	81.8	100

Based on table, 1 to define which classifier that has the best performance after implementing MiMaLo as follow:

a. Recall

Recall is the ability of the system to retrieve the information/data. The Decision Tree (DT) was able to get as 89.58%, k-Nearest Neighbor (kNN) was 93.33%, while Logistic Regression (LR) found 99.58%, Naïve Bayes (NB) only retrieved 88.75%, Neural Network (NN) was 97.50% and Support Vector Machine successfully achieved 100.00%. From these six classifiers, there is only one classifier that has the highest ability of recalling information is Support Vector Machine.

b. Precision

Precision is the accuracy in predicting the information in the correct way. Based on table 4.1, Decision Tree got 75.42%, k-Nearest Neighbor was 87.73%, Logistic Regression was 84.20%, Naïve Bayes was only 68.32%, Neural Network reached 90.51%, and Support Vector Machine hit 81.80%. Among these six Classifiers, Neural Network has the highest precision value.

c. Accuracy

Accuracy is a value of right prediction compared to the whole information or data. Decision Tree got 79.79%, k-Nearest Neighbor was 90.00%, Logistic Regression reached the second place with 90.21%, Naive Bayes was the lowest value only 75.21, Neural Network was the highest with value 93.54%, and Support Vector Machine was 88.54%.

d. ROC Curve

Fig. 2 shows the comparison of ROC Curve six classifiers of MiMaLo Algorithm. Referring to the performance in each classifiers that implemented MiMaLo algorithm and measuring the recall value, precision, accuracy, and Area Under Curve, thus, the best method in detecting malware with system call base using data mining classification technique is Neural Network with value of recall is 99.58%, precision value is 90.51%, accuracy reached 93.54%, and AUC as 0.982, is included in excellent category.

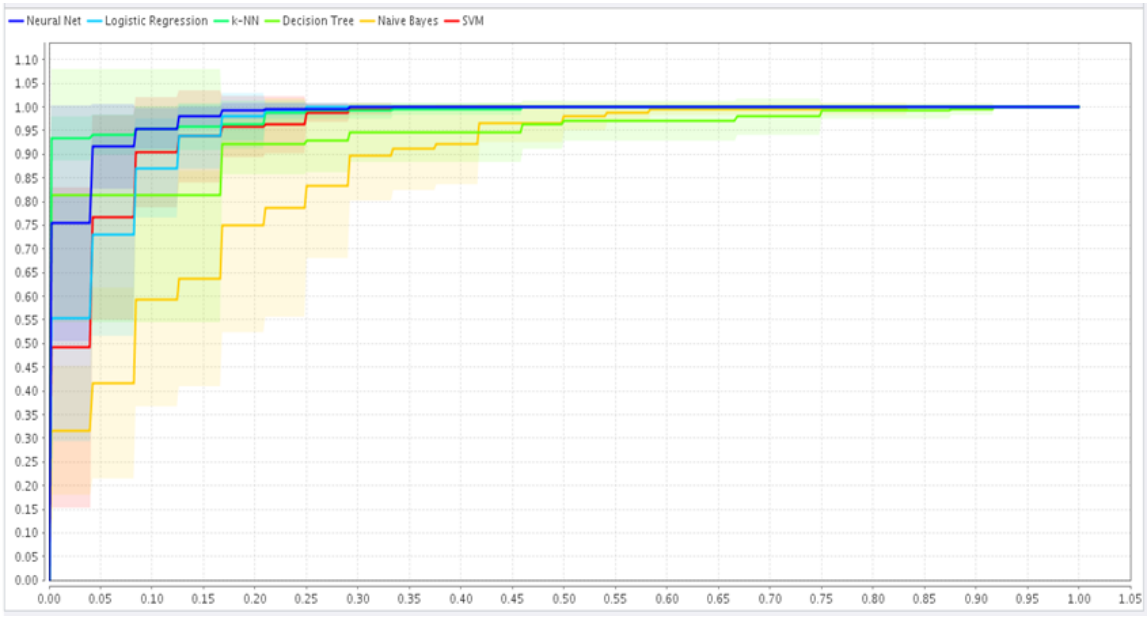

Fig. 2. ROC Comparison of MiMaLo

4.2. Performance Comparison: Basic Dataset vs Proposed Methods MiMaLo

Called from the result of the experiments of table 1 and table 2, each performance was compared as follows: Derived from table 2.

Table 2. Method Performance

#	CLASSIFIERS	ACCURACY	AUC	PRECISION	RECALL
1	DT	80	0.793	75.62	89.58
2	kNN	85	0.831	86.08	85
3	LR	86.25	0.92	87.28	85.42
4	NB	74.58	0.788	70.52	87.5
5	NN	81.67	0.931	80.21	87.08
6	SVM	79.38	0.894	71.77	97.5

The average of recall value from existing classifiers is 88.68%, the average value of precision is 78.58 %, the accuracy value is 81.14%, and the AUC is 0.859. For the proposed method "MiMaLo", based on able 4.2 , the mean value of recall is 94.65%, Precision is 80.68%, Accuracy is 85.62% , and AUC mean value is 0.900.

From the six classifiers used in the experiments, one of them experienced the declined but still in the same level, which is Decision Tree as shown in table 3.

Table 3. Performance Comparison-1

CLASSIFIERS	METHOD	ACCURACY	AUC	PRECISION	RECALL
DT	-	80	0.793	75.62	89.58
DT	MiMaLo	79.79	0.789	75.42	89.58

While the other five (5) classifiers were experiencing the significant rose as shown in table 4.

Table 4. Comparison Performance-2

#	CLASSIFIERS	METHOD	ACCURACY	AUC	PRECISION	RECALL
1	k-NN	-	85	0.831	86.08	85
	k-NN	MiMaLo	89.79	0.852	87.45	93.33
2	LR	-	86.25	0.92	87.28	85.42
	LR	MiMaLo	90	0.961	83.92	99.58
3	NB	-	74.58	0.788	70.52	87.5
	NB	MiMaLo	75.21	0.866	68.32	88.75
4	NN	-	81.67	0.931	80.21	87.08
	NN	MiMaLo	92.71	0.983	89.66	97.08
5	SVM	-	79.38	0.894	71.77	97.5
	SVM	MiMaLo	87.5	0.959	80.63	99.58

From these calculations, in summary, MiMaLo has been proven able to improve the performance of classifiers.

5. Conclusion

The proposed method, MiMaLo, successfully increasing the performance five classifiers, which are, k-Nearest Neighbor, Naïve Bayes, Logistic Regression, Support Vector Machine, and Neural Network, even though, one classifier, Decision Tree, its performance still at the same level.

Acknowledgements

Author thanks Rector of University of Technical Malaysia Melaka and Rector of Informatics and Business Institute Darmajaya Indonesia.

References

[1] Dimjasevic, M., Atzeni, S., Ugrina, I., & Rakamaric, Z. 2015. Android Malware Detection Based on System Calls. UUCS-15-003, 11(1), 209–216

[2] Comput, J.P.D., Tong, F., and Yan, Z., 2017. A Hybrid Approach Of Mobile Malware Detection In Android. J. Parallel Distrib. Comput., 103, pp.22–31.

[3] T.Bell.1999. The Concept of Dynamic Analysis. ACM SIGSOFT Softw. Eng. Notes.24, 6 (1999), 216-234.

[4] Lin, C.H., Pao, H.K., and Liao, J.W., 2018. Efficient Dynamic Malware Analysis Using Virtual Time Control Mechanics. Computers and Security, 73, pp.359–373.

[5] Or-Meir, O., Nissim, N., Elovici, Y., Rokach, L. 2019. Dynamic Malware Analysis in the Modern Era-A State of the Art Survey. ACM Computing Surveys, Vol.52, No.5, Articles 88. September 2019.

[6] Abela, K. J., Alas, J. R. D., Angeles, D. K., Tolentino, R. J., and Gomez, M. A., 2013. Automated Malware Detection for Android AMDA. In The Second International Conference on Cyber Security, Cyber Peacefare and Digital Forensic (CyberSec2013) pp. 180-188.

[7] Seo, S.-H., Gupta, A., Mohamed Sallam, A., Bertino, E., and Yim, K., 2014. Detecting Mobile Malware Threats To Homeland Security Through Static Analysis. Journal of Network and Computer Applications, 38, pp.43–53.

[8] Arp, D., Spreitzenbarth, M., Hübner, M., Gascon, H., and Rieck, K. 2014. Drebin: Efficient and Explainable Detection of Android Malware in Your Pocket. Proceeding of 17th Network and Distributed System Security Symposium (NDSS).

[9] Kabakus, A.T. and Dogru, I.A., 2018. An In-Depth Analysis Of Android Malware Using Hybrid Techniques. Digital Investigation, 24, pp.25–33.

[10] Faruki, P., Ganmoor, V., Laxmi, V., Gaur, M. S., and Bharmal, A., 2013. AndroSimilar: robust statistical feature signature for Android malware detection. In Proceedings of the 6th International Conference on Security of Information and Networks ACM, pp. 152-159.

[11] Lin, C.H., Pao, H. K. and Liao, J.W., 2018. Efficient dynamic malware analysis using virtual time control mechanics. Computers and Security, 73, pp. 359–373.

[12] Dini, G., Martinelli, F., Saracino, A., and Sgandurra, D. 2013. Probabilistic Contract Compliance for Mobile Applications. In Availability, Reliability and Security (ARES), 2013 Eighth International Conference on IEEE, pp. 599-606.

[13] Enck, W., Gilbert, P., Chun, B. G., Cox, L. P., Jung, J., McDaniel, P., and Sheth, A., 2010. TaintDroid: An Information-Flow Tracking System for Realtime Privacy Monitoring on Smartphones. In OSDI, 10, pp. 1-6.

[14] Xu, R., Saïdi, H. and Anderson, R., 2012. Aurasium: Practical policy enforcement for android applications. In Proceedings of the 21st USENIX conference on Security symposium. pp. 27-27.

[15] Wei, T. E., Mao, C. H., Jeng, A. B., Lee, H. M., Wang, H. T. and Wu, D. J., 2012. Android Malware Detection via a Latent Network Behavior Analysis. In Trust, Security and Privacy in Computing and Communications (TrustCom), 2012 IEEE 11th International Conference, pp. 1251-1258.

[16] Sanz B., Santos I., Ugarte-P. X., Laorden C., Nieves J. and Bringas P. G., 2013. Instance-based Anomaly Method for Android Malware Detection. In Proceedings of the 10th International Conference on Security and Cryptography (SECRYPT), pp. 387-394

[17] Jiawei, H., Kamber, M., Han, J., Kamber, M. and Pei, J., 2012. Data Mining: Concepts and Techniques. San Francisco, CA, itd: Morgan Kaufmann.

[18] Bebu, I., Luta, G., Mathew, T., Agan, K, B,. 2016. Generalized Confidence Intervals and Fiducial Intervals for Some Epidemiological Measures. International Journal of Environmental Research and Public Health. MPDI. 13 , 605. doi : 10.3390/ijerph13060605

Authors' Profiles

Sriyanto is currently a PhD student at the Universiti Teknikal Malaysia Melaka, Malaysia. He holds Bachelor of Computer Science from Gunadarma University, Indonesia and a master degree in Information System Management from Gunadarma University, Indonesia. His research interests include Intrusion Detection System, Artificial Intelligence and Security System.

Shahrin Sahib received the Bachelor of Science in Engineering, Computer Systems and Master of Science in Engineering, System Software in Purdue University in 1989 and 1991 respectively. He received the Doctor of Philosophy, Parallel Processing from University of Sheffield in 1995. He is a professor of Faculty of Information Technology and Communication at the Universiti Teknikal Malaysia Melaka. His research interests include network security, computer system security, network administration and network design. He is a member panel of Experts National ICT Security and Emergency Response Center and also Member of Technical Working Group: Policy.

Assoc. Prof. Dr. Mohd Faizal Abdollah is currently a senior lecturer in University Teknikal Malaysia Melaka. The research area more focuses on network security, malware detection and network management. In cybersecurity, Dr Mohd Faizal led the sub project under CMERP project with the collaboration with Cyber Security Malaysia. This project more focuses on malware detection, eradication and mitigation. Others than that, Dr Mohd Faizal also involve in various grant sponsor by Ministry of Education, Industrial grant and University grant such as Fundamental Grant for detecting botnet activity, Transdiciplin Grant for detecting the inside threat, ISIF grant for botnet detection using graph theory. He also teaches UTeM course such as Information Technology and IT Security, Network Management and Administration, Advanced Scalable Network and also manage to produce various conference paper and journal in cybersecurity related field.

Nanna Suryana is currently a Senior Lecturer and professor at the University of Malaysia, Kuala Lumpur, Malaysia. His current research interest includes: Artificial Intelligence, Software Engineering, Information System, Data Mining and Image Processing.

Adang Suhendra is currently a senior lecturer expertise in Computer Graphics, Computer Vision and Software Engineering at the Gunadarma University, Indonesia. His current research interest includes Development of Talking Avatar System, Computer Graphics Modelling and Animation, Driving Simulator and Virtual Crane Simulator.

How to cite this paper: Sriyanto, Sahib B. Sahrin, Abdullah Mohd. Faizal, Nanna Suryana, Adang Suhendra, "MiMaLo: Advanced Normalization Method for Mobile Malware Detection", International Journal of Modern Education and Computer Science(IJMECS), Vol.14, No.5, pp. 24-33, 2022. DOI:10.5815/ijmecs.2022.05.03