Journal of Information Technology and Digital World (2019)
Vol.01/ No. 02
Pages: 96-107
https://www.irojournals.com/itdw/
DOI: https://doi.org/10.36548/jitdw.2019.2.005

MACHINE LEARNING BASED AUTOMATIC PERMISSION GRANTING AND MALWARE IDENTIFICATION

Dr. M. Duraipandian,
Head of Department, Department of Computer Science and Engineering,
Vivekanandha College of Technology for Women,
Namakal, India.
Email: svsduraipandian@gmail.com

Mr. R. Vinothkanna,
Department of ECE, Vivekanandha College of Technology for Women,
Namakal, India.
Email: rvinothkannaphd@gmail.com

Abstract: The mobile device have gained an imperative predominance in the daily routine of our lives, by keeping us connected to the real world seamlessly. Most of the mobile devices are built on android whose security mechanism is totally permission based controlling the applications from accessing the core details of the devices and the users. Even after understanding the permission system often the mobile user are ignorant about the common threat, due to the applications popularity and proceed with the installation process not aware of the targets of the application developer. The aim of the paper is to devise malware detection with the automatic permission granting employing the machine learning techniques. The different machine learning methods are engaged in the malware detection and analyzed. The results are observed to note down the approaches that aids better in enhancing the user awareness and reducing the malware threats, by detecting the malwares of the applications.

Keywords: Machine Learning, Malware Detection, Mobile Devices, Android Application, Permission Granting

1. INTRODUCTION

Recently the uses and the production of the mobile phones are progressing at a rapid pace, along with the technology that aims to gain processors with the high performance, storage enhancements, low power consumption, high speed internet connection and touch screen with high resolution. Apart from this the mobile application with the rich content enables the certain sophisticated actions, occupying a predominance space in the life of humans.

The android operating system that holds the highest market share as on 2014 is a very common operating system found in the mobile phones. The smart phone enabled with the android also prevails as a prime target for the hackers as it is an open source operating system powered by the Linux platform. The security mechanism of the malware is designed

ISSN: 2582-418X (online)

Journal of Information Technology and Digital World (2019)
Vol.01/ No. 02
Pages: 96-107
https://www.irojournals.com/itdw/
DOI: https://doi.org/10.36548/jitdw.2019.2.005

inform the users about the permission requested by the application and install the application once the permission is granted. But in most cases the user fails to see the permission that is requested as the application seem to be more attractive and the user does not have enough standards or the level to understand the intentions of the application provider

For example the terms and the conditions agreed by the users , only 10%user takes the pain to read and agree, whereas all the other just agree without knowing what are the permission requested, even legitimate applications are altered injecting malicious codes to hack the personal details of the users. One of the very common permission accesses is the acceptance of the cookies while accessing the information through the internet.

So the paper aims in devising a automatic malware detection system that is based on permission granting, by developing machine learning models based on the behavior of the applications and the pattern followed in the permission further the machine learning is also employed to analyze the source code to classify the malicious injected codes in the legitimate applications and the malicious applications. The rest of the paper is arranged as 2. Related works, 3. Proposed work and 4. Results evaluation and 5.Conclusion

2. RELATED WORKS

Talha, Kabakus et al [1] proposes the "APK auditor that is comprised of signature database, android client and the central server to analyze and store the information of the app, to grant application analysis request and to extend communication with the signature, database and mobile user managing the complete process"

Qiao, Mengyu et al [2] the proposed method in the paper puts into action the statistical approaches to detect the malware in the by mining the patterns of the permission as well as the application interface function calls that are acquired and used by the android. Mariconti et al [3] the author presents an "android malware detection system that is completely relies on the behavior of the application, this is coined as MAMADROID. The behavior of the application is developed as a Markov chain from the sequence of the abstracted application interface calls. The table .1 below provides the profiling of the android application.

ISSN: 2582-418X (online)

Journal of Information Technology and Digital World (2019)
Vol.01/ No. 02
Pages: 96-107
https://www.irojournals.com/itdw/
DOI: https://doi.org/10.36548/jitdw.2019.2.005

App Profiling	Description
App components	Four different components defined in the android app are activity, service, content provider , broadcast receiver
Intents	this is responsible for communicating between apps and different components of an app.
Requested Permission	Permissions are actively granted by the user when an app is installed, and they are mainly used to limit the use of some functions and the access to some components among apps
Hardware	If an app needs to request the camera, Bluetooth, or the GPS module of the smartphone, these features have to be declared in the manifest file. Besides, if an app requests GPS and network at the same time, then it means that the app can read location information and reveal it out through network
API calls	Set of functions provided to control principal actions of Android OS.
Protected strings	Analyzed malware samples and some defined suspicious strings
Commands	Android OS is developed based on Linux and uses some commands as Linux.
Network	Android apps require network access during running time

Table .1 Android Application Profiling

Wang, et al [4] employs three machine learning approaches in parallel KNN, random forest and J48 and does information fusion using either the probabilistic analysis based fusion, dempster-shafer theory based fusion, to detect the malware in the android applications. Wu, et al [5] the "droiddolphin is proposed to detect the malicious applications of the android by leveraging the technologies of GUI based testing, big data analysis framework and the machine learning.

Chen et al [6] puts forth the machine learning based malware detection StormDroid to support scalable, malware detection methodologies observing the application behavior statistically and dynamically. Liu, Xing, et al [7] the author has proposed two layered permission based malware detection as the defense and control mechanism against the malicious applications.

Joseph, et al [8] presents the intelligent computing enabled by the data mining as the survey, Arp, et al [9] the paper proposes a light weight malware detection method to identify the malicious attacks performing a broad static analysis by collecting as many applications as possible. Smys, S. et al [10] proposes the machine learning approaches to detect

ISSN: 2582-418X (online)

Journal of Information Technology and Digital World (2019)
Vol.01/ No. 02
Pages: 96-107
https://www.irojournals.com/itdw/
DOI: https://doi.org/10.36548/jitdw.2019.2.005

the Distributed denial of service in the telecommunication networks. Sathesh, et al [11] proposes the methodology of intrusion detection and applies them to detect the intrusions in the social network.

Rovelli, eta l [12] has put forth the method to automatically detect the unseen harmful behavior, utilizing the machine classifier that is constructed based on the behavioral markers. G. Josemin Bala et al [13] proposes a "Construction of virtual backbone to support mobility in MANET—A less overhead approach." .Yang, et al [14] "Droidminer: Automated mining and characterization of fine-grained malicious behaviors in android applications."

Raj, Jennifer S et al [15] presents the "survey on the computational intelligence techniques and its applications." Feldman et al [16] "Manilyzer: automated android malware detection through manifest analysis." Moonsamy et al [17] "Mining permission patterns for contrasting clean and malicious android applications." Anguraj, et al [18] ""Trust-based intrusion detection and clustering approach for wireless body area networks." Praveena, A et al [19] has put forth the cryptographic approaches.

3. THE SYSTEM MODEL FOR PROPOSED MALWARE DETECTION

Though the android has its own security mechanism, the android devices are still vulnerable to the malwares, the recent reports on the android malware samples shows the exponential growth of the malware at a rapid pace, the number of malwares prevailing now is about 14 times of that which existed in the year 2011. The types of malwares that were identified are data stealers that procure all the personal information of the user from the mobile phone and send it to third party, rooting capable, that takes complete control of the mobile devices and functions, and encourages self-replication without revealing its identity, Premium service abusers, mobile device spies etc. and android has four level of permissions that are based on the intended use. They are the normal permission, dangerous permission signature permissions and system permissions [2]. The conventional machine learning algorithms such as the decision tree and the SVM learn one hypothesis from the training data, but these classification algorithms do not work well real world android malware. So the proposed method aims in devising the automatic malware detection system that is based on permission granting, by developing machine learning models based on the behavior of the applications and the pattern followed in the permission further the machine learning is also employed to analyze the source code to classify the malicious injected codes in the legitimate applications and the malicious applications. The figure.1 below shows the flow diagram of the proposed malware detection machine learning model to classify the malware applications from the genuine.

ISSN: 2582-418X (online)

Journal of Information Technology and Digital World (2019)
Vol.01/ No. 02
Pages: 96-107
https://www.irojournals.com/itdw/
DOI: https://doi.org/10.36548/jitdw.2019.2.005

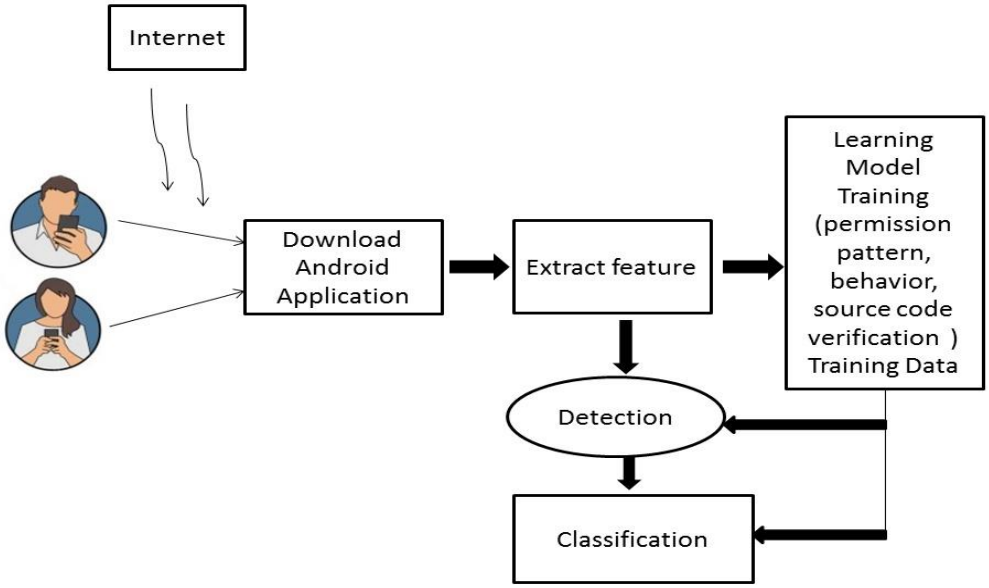


Figure. 1 Proposed Flow Diagrams

3.1. METHODOLOGY

Whenever the user tries to download an application from the Google play store the application requests for the permission, and the proposed method, extracts the features and the applications such as its behavior, permission pattern and the source code and trains the machine learning model (MLM-Droid) on the patterns of the permission the behavior of the application and source code and the trained data are fed to the detection and the classification process to identify the malicious applications that can affect the performance. The MLM-Droid uses the logistic regression, CART, Naïve Bayes and the random forest to classify the malicious with the benign. Some of the common malicious patterns are that which request the access of contact details, location along with the camera accessibility, sign in procedures using the existing Google accounts and access to personal information, in countries like India some applications request for the permanent account number to proceed further in installing an application for e.g. the android applications related to Jobs the some of the permission patterns are listed below in the table.2. The list of permissions that were used and that were requested were gathered applying the clustering method [3] to generalize the permission patterns that are most often requested and the source code for the permission declaration was also gathered and fed to the training model, to identify the changes injected to the source code. The behavior, pattern and the source code of the permission were used as training data to the MLM –Droid to classify between the applications

ISSN: 2582-418X (online)

Journal of Information Technology and Digital World (2019)
Vol.01/ No. 02
Pages: 96-107
https://www.irojournals.com/itdw/
DOI: https://doi.org/10.36548/jitdw.2019.2.005

that are malicious from the benign. The figure.2 shows the methodology adopted to classify the malicious from the genuine

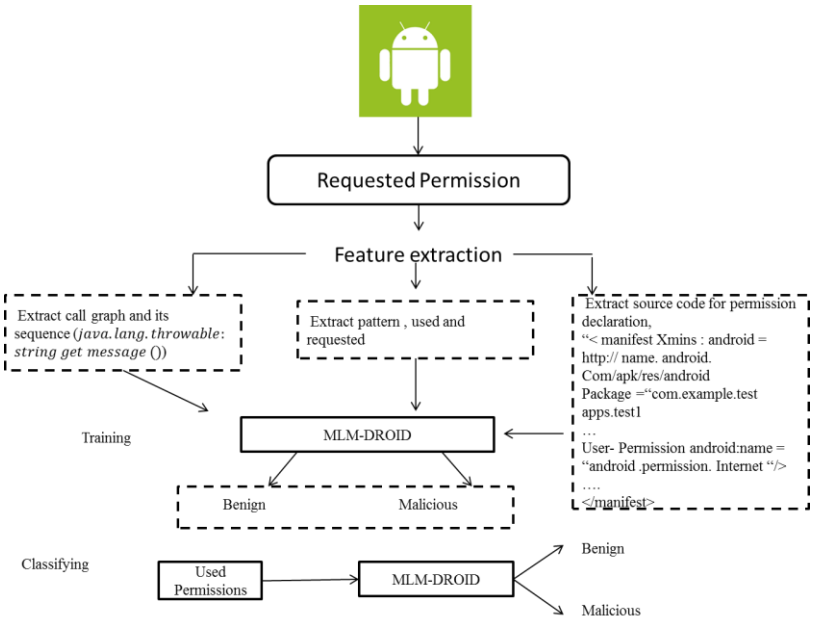


Figure.2 Proposed Methodology

The samples of the used and the requested permissions the frequency of their occurrence are listed below in the table.2 The MLM-Droid uses the multiple machine learning algorithms such as the logistic regression, CART, Naïve Bayes and the random forest in parallel to classify the malicious and the genuine.

ISSN: 2582-418X (online)

Journal of Information Technology and Digital World (2019)
Vol.01/ No. 02
Pages: 96-107
https://www.irojournals.com/itdw/
DOI: https://doi.org/10.36548/jitdw.2019.2.005

Permission	Types	Description	Frequency %
Requested	Genuine (benign)	Internet	91.36
		Network Access	55
		Vibrate	20
		Location Access	20
		Phone State Accessing	31
	Malicious	Internet	97.72
		Network Access	93
		Vibrate	81
		Location Access	95
		Phone State Accessing	33
Used	Genuine (benign)	Internet	83
		Network Access	60
		Vibrate	49
		Location Access	30
		Phone State Accessing	37
	Malicious	Internet	95
		Network Access	42
		Vibrate	77
		Location Access	16
		Phone State Accessing	38

Table.2 Samples of used and Requested Permissions

The table represents few samples of the permissions that are common to both the requested and the used; there are also other unique permissions that are requested and used some of them are Expand-status bar, read history book marks, account detail access etc. The MLM-Droid analyzes automatically the feature vectors of all the applications and classifies the genuine from the malicious; it engages multiple machine learning algorithms in parallel implementing the algorithms using the Weka open source suite machine learning tool. In addition to classification the trends of the permissions are also analyzed.

ISSN: 2582-418X (online)

Journal of Information Technology and Digital World (2019)
Vol.01/ No. 02
Pages: 96-107
https://www.irojournals.com/itdw/
DOI: https://doi.org/10.36548/jitdw.2019.2.005

4. RESULT ANALYSIS

The efficacy of the MLM-Droid is estimated by examining the machine learning algorithm's engaged in proposed method, the competence of the system is tested applying the formulas, of accuracy, precision, False positive rate and the true positive rate, the proposed method is compared with the existing methods such as StormDroid [6], MAMA Droid [3] Droid dolphin [5] and the Mlifdect [4] on the terms of detection rate, accuracy, recall, precisions and f-score.

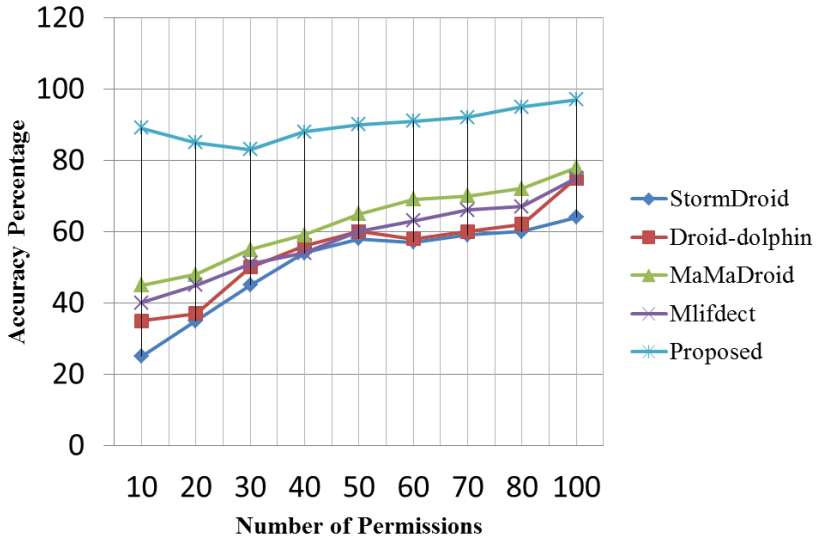


Figure.3 Accuracy Observed

The figure.3 shows the accuracy achieved for the number of permissions for the proposed and the existing methods. The results observed in the figure.3 provides the percentage of accuracy attained by the proposed and the existing methods, The accuracy percentage of the proposed method is 25% higher compared to the StormDroid , 15 % better than the MAMA Droid 21% higher than Droid dolphin and 13% higher than the Mlifdect.

ISSN: 2582-418X (online)

Journal of Information Technology and Digital World (2019)
Vol.01/ No. 02
Pages: 96-107
https://www.irojournals.com/itdw/
DOI: https://doi.org/10.36548/jitdw.2019.2.005

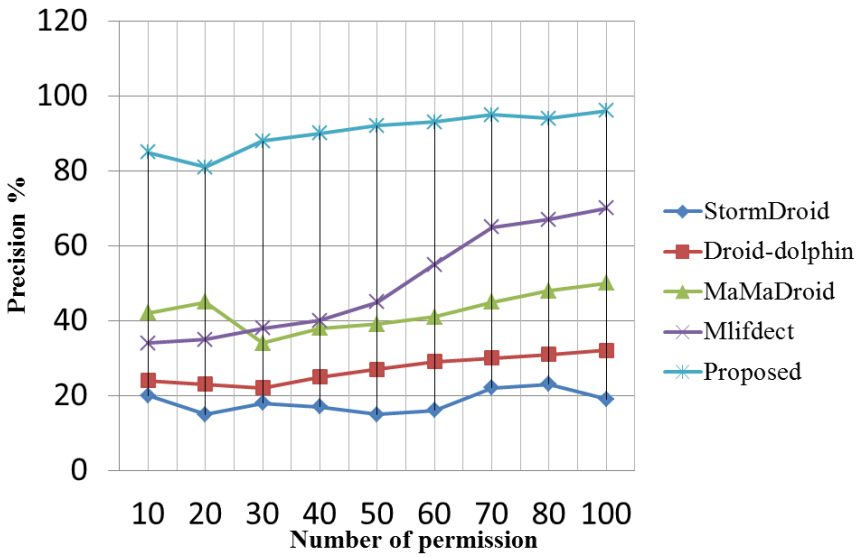
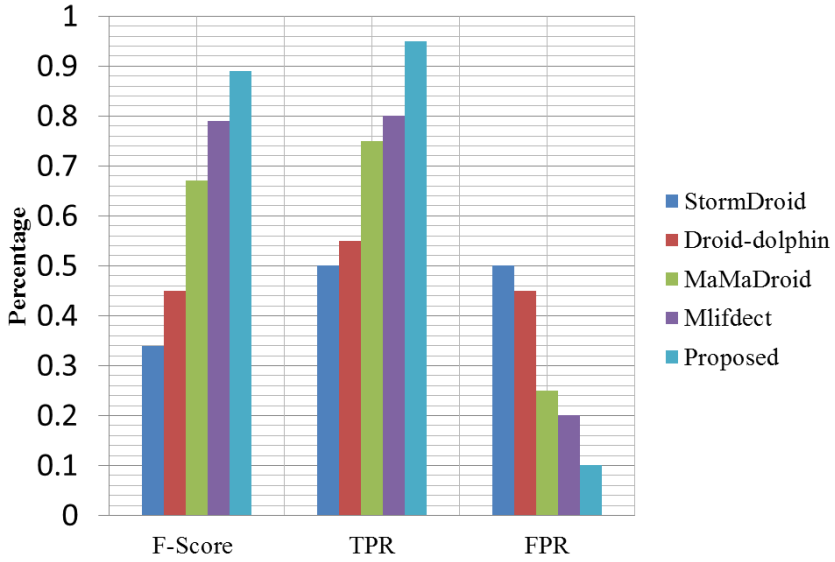


Figure.4 Precision observed

The figure.4 above shows the percentage of precision observed for the proposed and the existing methods, precision percentage of the proposed method is 10% higher compared to the StormDroid, 12 % better than the MAMA Droid 8% higher than Droid dolphin and 13% higher than the Mlifdect.



ISSN: 2582-418X (online)

Journal of Information Technology and Digital World (2019)
Vol.01/ No. 02
Pages: 96-107
https://www.irojournals.com/itdw/
DOI: https://doi.org/10.36548/jitdw.2019.2.005

Figure.5 F-score, TPR, FPR

The figure.5 above presents the observed F-score, true positive rate (TPR) and the false positive rate (FPR) of the proposed and the existing methods StormDroid, MAMA Droid, Droid dolphin and Mlifdect and the figure.6 below provides the actual number of the genuine and malicious apps based on the permissions.

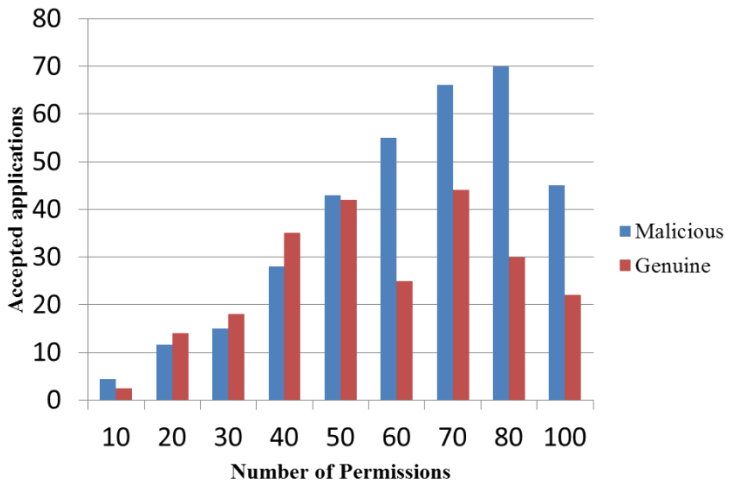


Figure.6 Accepted Applications

5. CONCLUSION

The paper focusing on identifying the malicious android applications in the mobile devices, has devised a machine learning model (MLM-Droid) employing multiple machine learning approaches in parallel to classify the genuine and the dangerous application. The features of the application such as the source code of the permission, the behavior and the pattern are gathered and used for training. The trained data is used to detect and classify the applications nature. The results observed shows that the proposed method has enhanced accuracy, precision, F-score, true positive rate and false positive rate. In future the paper is to apply the deep learning methods in identifying the genuineness of the mobile applications.

References

ISSN: 2582-418X (online)

Journal of Information Technology and Digital World (2019)
Vol.01/ No. 02
Pages: 96-107
https://www.irojournals.com/itdw/
DOI: https://doi.org/10.36548/jitdw.2019.2.005

[1] Talha, Kabakus Abdullah, Dogru Ibrahim Alper, and Cetin Aydin. "APK Auditor: Permission-based Android malware detection system." *Digital Investigation* 13 (2015): 1-14.

[2] Qiao, Mengyu, Andrew H. Sung, and Qingzhong Liu. "Merging permission and API features for Android malware detection." In *2016 5th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*, pp. 566-571. IEEE, 2016.

[3] Mariconti, Enrico, Lucky Onwuzurike, Panagiotis Andriotis, Emiliano De Cristofaro, Gordon Ross, and Gianluca Stringhini. "Mamadroid: Detecting android malware by building markov chains of behavioral models." *arXiv preprint arXiv:1612.04433* (2016).

[4] Wang, Xin, Dafang Zhang, Xin Su, and Wenjia Li. "Mlifdect: android malware detection based on parallel machine learning and information fusion." *Security and Communication Networks* 2017 (2017).

[5] Wu, Wen-Chieh, and Shih-Hao Hung. "DroidDolphin: a dynamic Android malware detection framework using big data and machine learning." In *Proceedings of the 2014 Conference on Research in Adaptive and Convergent Systems*, pp. 247-252. 2014.

[6] Chen, Sen, Minhui Xue, Zhushou Tang, Lihua Xu, and Haojin Zhu. "Stormdroid: A streaminglized machine learning-based system for detecting android malware." In *Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security*, pp. 377-388. 2016.

[7] Liu, Xing, and Jiqiang Liu. "A two-layered permission-based android malware detection scheme." In *2014 2nd IEEE International Conference on Mobile Cloud Computing, Services, and Engineering*, pp. 142-148. IEEE, 2014.

[8] Joseph, S. I. T. (2019). SURVEY OF DATA MINING ALGORITHM'S FOR INTELLIGENT COMPUTING SYSTEM. Journal of trends in Computer Science and Smart technology (TCSST), 1(01), 14-24.

[9] Arp, Daniel, Michael Spreitzenbarth, Malte Hubner, Hugo Gascon, Konrad Rieck, and C. E. R. T. Siemens. "Drebin: Effective and explainable detection of android malware in your pocket." In *Ndss*, vol. 14, pp. 23-26. 2014.

[10] Smys, S. (2019). DDOS ATTACK DETECTION IN TELECOMMUNICATION NETWORK USING MACHINE LEARNING. Journal of Ubiquitous Computing and Communication Technologies (UCCT), 1(01), 33-44.

[11] Sathesh, A. (2019). ENHANCED SOFT COMPUTING APPROACHES FOR INTRUSION DETECTION SCHEMES IN SOCIAL MEDIA NETWORKS. Journal of Soft Computing Paradigm (JSCP), 1(02), 69-79.

[12] Rovelli, Paolo, and Ýmir Vigfússon. "Pmds: Permission-based malware detection system." In *International Conference on Information Systems Security*, pp. 338-357. Springer, Cham, 2014.

ISSN: 2582-418X (online)

[13] Smys, S., G. Josemin Bala, and Jennifer S. Raj. "Construction of virtual backbone to support mobility in MANET—A less overhead approach." In *2009 international conference on application of information and communication technologies*, pp. 1-4. IEEE, 2009

[14] .Yang, Chao, Zhaoyan Xu, Guofei Gu, Vinod Yegneswaran, and Phillip Porras. "Droidminer: Automated mining and characterization of fine-grained malicious behaviors in android applications." In *European symposium on research in computer security*, pp. 163-182. Springer, Cham, 2014.

[15] Raj, Jennifer S. "A COMPREHENSIVE SURVEY ON THE COMPUTATIONAL INTELLIGENCE TECHNIQUES AND ITS APPLICATIONS." Journal of ISMAC 1, no. 03 (2019): 147-159.

[16] Feldman, Stephen, Dillon Stadther, and Bing Wang. "Manilyzer: automated android malware detection through manifest analysis." In *2014 IEEE 11th International Conference on Mobile Ad Hoc and Sensor Systems*, pp. 767-772. IEEE, 2014.

[17] Moonsamy, Veelasha, Jia Rong, and Shaowu Liu. "Mining permission patterns for contrasting clean and malicious android applications." *Future Generation Computer Systems* 36 (2014): 122-132.

[18] Anguraj, Dinesh Kumar, and S. Smys. "Trust-based intrusion detection and clustering approach for wireless body area networks." *Wireless Personal Communications* 104, no. 1 (2019): 1-20..

[19] Praveena, A., and S. Smys. "Efficient cryptographic approach for data security in wireless sensor networks using MES VU." In *2016 10th international conference on intelligent systems and control (ISCO)*, pp. 1-6. IEEE, 2016.

ISSN: 2582-418X (online)