Dynamic Malware Analysis Using Machine Learning Algorithm

Udayakumar N, Anandaselvi S, Subbulakshmi T
School of Computing Science and Engineering
VIT University Chennai
Chennai, India
udayakumar.n2014@vit.ac.in, sanandaselvi.2017@vitstudent.ac.in, research.subbulakshmi@gmail.com

Abstract— **Malware detection is a vital think about the protection of the Personal computer systems. However, presently using signature-based strategies cannot offer correct detection of zero-day attacks and polymorphic viruses. That's why the requirement for machine learning-based detection arises.**

The purpose of this work was to work out the most effective feature extraction, feature illustration, and classification ways that end in the most effective accuracy. This work presents suggested ways for machine learning based malware classification and detection, also as the tips for its implementation. Moreover, the study performed is often helpful as a base for any analysis within the field of malware analysis with machine learning strategies

Keywords— Dynamic malware analysis, machine learning

I. INTRODUCTION

Dynamic malware analysis is executed based on the behaviour of the malware. Behaviour analysis gives the better understanding of the malware i.e. When it get installed and who talks to it information can be gained for dynamic malware analysis safe environment need to be set up. In this is the unrestricted program are securely running in a safe environment are monitored Here two snapshot are taken for comparison Active process are listed out Process registry are monitored Internet service is stimulated Check whether them malware is connected to the internet or not

II. LITERATUR SURVEY

Zongqu Zhao[1]proposed a manual analysis to extract the signature of the malware(It contain 9398 PE files that include 4828 malware) can be but it is inefficient and there is a limitation on detecting unknown one , when they deal with large amount of malware ,proposed work involve malware cross reference and feature selection approach. Proposed work can extract opcode sequence and translate them into features by vector space model. Data mining algorithm are employed to find the classify rules for detecting malware. Syntactic structure of the code is not considered when they extract the byte sequence from the software there is lack of compromise between the universality and discrimination of the feature is the main problem. The labelled files are disassembled and opcode chosen and segmented based on control flow .using Boolean vector the feature is obtained to detect the executable is malware or not .Author conclude that contribution of this project is they detected the unknown and known malware and also they achieved good performance in detection rate (overall 97% is achieved)using control-flow construct.

Chang-zhen Hu[2] proposed the malware variants(472 Linux malware samples collected) and the Maliciousness of a program can be determined using malware identification .some technique have been developed to detect rapid expansion of malware but they did not detect malware instances in proposed work novel dynamic malware method is used it contain multiple execution path and trigger malicious behaviour In this MBO is constructed to identify malware family and vm is developed to trigger mbo . Triggered malicious behaviours are modelled which can identify unknown malware variants. Problem in this paper is 1) using the same clock interrupt and the i/o exception may produce the risk in exposing the behaviour.2)The second limitation is anti-virtualisation techniques can detect virtual machine and debugger. Implementation done using 1)MBO is used detect malicious features and behaviour ,advantages is that false positive rate is reduced and alert is given if we detect malicious behaviour along with resulted outcome2)QEMU(tool) is used to develop virtual monitor the task of virtual monitor is to emulate virtual monitor 3)In multipath exploration the Breadth - first traversal is used to improve the speed of analysis. The author concludes that A novel malware identifier is the proposed work to detect. Malicious feature are established using mbo. Malicious feature matrix is build based on the detected malicious behaviours in the virtual monitor and based on the MBO .higher rate is achieved by determining multipath exploration depth in proposed work

Sanjeev Das[3] proposed that Conventional software is not successful in defending malicious program so in this paper the guard ol approach is a combined approach with field-programmable gate array (FPGA) is used to detect online malware. In Syscall extractor by executing the program system call and argument return value is extracted. The Classifier trainer trains the classifier engine from the constructed feature vector. Linux OS running on a 32-bit Intel x86 processor In this hardware enhanced architecture guard is used to detect malware at run time .I extract system call and semantic of malicious behaviour and novel frequency-centralized model is used to extract the feature which is obtained from the malware sample .the result show that guard ol is fast and effective

Fei Zhang[4] proposed that adversarial setting have been increasingly adopted in Pattern recognition and machine learning techniques. Adversary-aware classification algorithms are focused on previous work but reduced feature sets on classifier security against the same attacks are considered only by the few author .In this paper we focus on evasion attack and novel adversary-aware feature selection is proposed for improving the security against evasion attack. This paper provide understanding about the vulnerabilities of feature selection methods in adversarial settings, and paper aim towards developing more secure feature selection schemes against adversarial attacks

Implementation is done using 1)Wrapper-Based Adversarial Feature Selection, With Forward Selection (FS) and Backward Elimination (BE) 2)Evasion attack 3) security evaluation 4)spam filtering5) malware detection in pdf .In proposed work the generalization capability of the wrapped classifier, security against evasion attacks are optimized by adversarial feature selection method. In Complex feature mapping there is a direct relationship with the characteristics of the sample and it is not difficult to modify the malicious sample for finding optimal evasion point .this is why application –specific issue is the inverse feature mapping problem

GUODONG ZHAO1[5] proposed that Serious threat to the Internet is a Advanced persistent threat (APT. Command and control(C&C)servers is located using DNS.A novel system approach is proposed in this paper to detect APT malware infections based on malicious DNS and traffic analysis. APT malware C&C domains are detected using DNS analysis techniques. The volume of network traffic that needs to be recorded by security approaches it also improve the sustainability of the system

Problem: the limitation of ideas is does not good at detecting the malware that rely on the domain .1) Data (900 domains from VRT rule sets, malware samples of email attachment is collected) is collected from the collected data the network and DNS traffic is analysed. 2) from the DNS traffic the malicious DNS is detected and from the network traffic the signature based **DNS** is detected. For detecting the malware infection inside the network with DNS traffic, a novel ideas system is proposed. The proposed work is good detecting APT malware with high efficiency and accuracy.

E. Aharoni[6] proposed To identify suspicious activity they developed a system that continuously and automatically processes streaming data. That data contains low-level traces of process activity .if the system detect the malicious activity it give signal to inspection server.

This paper focus on both binary related feature and general event feature but result can be improved if it solely rely on binary related feature

Collect and process the event and retrain a new model based on the last 3 months of data Final model produces result whether the data is malware or not and provide the confidence score Database contain suspicion score and predictive information about the MD5.

X. Hu J. Jang T[7] proposed work solves the Microsoft Malware Classification Challenge on the Kaggle Platform the original space is high-dimensional, and the input data are linearly separable, and hence there is no need to map the input vectors to a higher-dimensional feature space. For PE files, hex dumps converted into raw binary to reconstruct executable PE files and by PE headers are reconstructed based on the information from the IDA Pro outputs Virus total are used for feature extraction it produce the result of 40 different antivirus product Malware is trained to identify malware family

Based on the machine instruction and AV malware classification can be done from the malware sample to identify the malware family

Xiaolin Gui,[8] Security and privacy is the critical issue where x-code is the recent malware and it is based on http request ,it is an ios malware lots of application is infected by xcode .in this they proposed a heuristic model based on fingerprint to and web is used to identify infected application. Limitation in the project is More than 60%of ios application are affected. The traffic characteristics of xcode is based onreal network traffic .During the initial exploration, it is found that more than 60% of iPhone were infected and the time and the traffic characteristic of x-code ghost were analysed .from the result we found that infected app send web-related info to the affected server

Suchul Lee[9] proposed method, called LDA-based Automatic Rule Generation (LARGen), automatically performs an analysis of the malicious traffic and extracts the appropriate attack signatures that will be used for IDS rules implementation is done using top 10% of signatures are chosen to create an IDS rule and be examined through the virtual IDS engine. uploaded to ids engine and evaluated , transferred to virtual engine from this author concludes that LARGen that automatically performs malicious traffic analysis and generates attack signatures. We first explored the effectiveness of LDA in the context of network attack signature generation. To elaborate the LDA-based automatic signature generation, we performed extensive experiments with real network trace data. Our experimental results revealed that the threat rules generated from LARGen accurately detect cyber-attack with most 1.6% false positives.

Wojciech Mazurczy,[10] proposed the information hiding is used to make data difficult to notice but it Is often neglected by security community ,but it is used to ex-filtrate data and make security threats. Steganography is one of the well-known fields.

Problem: a posteriori approach is very difficult because nor any universal countermeasures, Implementation is done using Conclusion information hiding is the serious threat but we does not make much effort to solve this.

Cai Fu,Xiao-Yang[11] Liu proposed a epidemic feedback model and quantitative analyse propagation effect based on density propagation effect threshold and reproduction number Using The real data sets CDBLP, Facebook, Weibo, and P2P analyse is done. From the experiment virtual virus pool is formed which made propagation easy and quick. Velocity and density of the virus is verified and result show that virus propagation in social network is accelerated using search engine

Domhnall Carlin [12] proposed that when the malware is reclassified with the labels and the detection is done using count based algorithm and sequence based algorithm but count based algorithm provide better result than sequence based algorithm

Matthew Tischer [13] Conducted an experiment with 300 USB flash drives among them 68% didn't took any precaution and only 16% scanned the drive and ,during the attack expected files are replaced with html file or it may contain Trojan horse ,which contain the embedded image without executing any code it tracks our work I.e. it may steal our sensitive data. Author concludes that this type of attack would affect the individual rather than the organisation and the author says that sometimes the simplest attack is more threatful than other attack

Yacin Nadji[14] proposed rza algorithm to measure the reason for botnet and and they focus on DNS and command and control servers .Data set comprises of passive dna and malicious library file .in this the single malware sample are kept under the different malware sample to learn the use of additional domain in c&C and then isolated to know the behaviour of running each sample .in this paper we study 3322.org NS to study the nitol botnet. Limitation of this paper is rza is focused on c&c and DGA .and the enumeration fail if the c&c does not share the IP address from this author concludes that rza is useful in helping both expedite the take down and the future take down process

Martin Courtney[15] stated that there is a significant rise in hacking and malware tools but the biggest problem is to detect the source of the attack and also the sophisticated attack is also increasing. But there is a slow development in cyber security vedorin designing the cyber security tools till now author conclude that we need to work to overcome the international threat by enhancing the tools

Zong-Xian Shen[16], Security Semantics Modelling with Progressive Distillation proposed that the conventional approach did not allow the user to scan the app before installing ,so this paper the petrdish method is proposed for repacked malicious app .this petridish method generates byte code to generate detection model and from the foot print it extract the malicious software proposed work used 1115 malicious program

Nayeem Islam[17] stated that Cyber-attacks can be launched by malware applications but have been downloaded from an app store or side-loaded to the device but that steal the information. And gain the access of the phone here the aggregated machine learning classifier ids used and both the behaviour and static analysis is used to learn the behaviour of the product .Wi-Fi also plays the major gate way for the major threat so too overcome this machine learning algorithm is used here.

Mingshen[18] Sun states that due to the popularity of android it also attract many malware 98% of new malware were from android so in our experiment they collected 2723 malware sample and achieved 99% accuracy in malware detection. This paper to represent the runtime behaviour of malware it has been detected by RBG and SSN behaviour of malware is also detected

Bo Liu,acc[19] stated that by applying the network capacities the bond of infection rate is analysed next the impact of mobility is studied relation between the different scheme is provided. Epidemic propagation model is used in first the malware propagation is compared using 4 different feature broad cast takes less time to infect the node and it is dangerous then then unicast and spread of the malware is more dangerous in broad cast

R. J. **Mangialardo[20]** and J. C. Duarte states that both the static and dynamic malware analysis has some disadvantage so in this method the author integrated both static and dynamic malware analysis by using random forest algorithm and FAMA frame work is used to classify the malware and the experiment provided better result

Andrea Saracino[21], states that android is more popular not only for its user friendly application but also for its malware attack and the user faces serious threat to their privacy to stop the malicious activity MADAM has been designed the dataset comprises of 2000 app from that 50 malware families are classified. Performance overhead has been measured in madem algorithm from 2011 the attacker have been increased threats in android is also increasing madem algorithm achieved 93% detection rate

Jemal Abawajy[22] states that android device has been facing a serious threat in malware attack due to its popularity .so this paper author propose Iterative Classifier Fusion System(ICFS) smallest number of classifier is used here but it produces best result using Libsvm with kernel weka toolkit is used here for implementing ICFS Malgenome data set is use here .Basic selection feature is performed here and triplet of the classifier is used here higher AUC is achieved in this performance

JUN-WON HO[23] states that in sensor network self-propagating malware is the serious threat worm propagation sensor field is used here randomized sprt scheme is proposed this technique achieve the fast propagation result we also evaluated the scheme through simulation

Ke Tian, Danfeng[24] proposed that states that to detect malware in android repacked technique is used along with partition based detection is used which reduces false positive rate and to detect the finger print fuzzy hacking technique is used.which produces better result compared to previous technique

Liang Xiao,[25] As accurate malware detection on mobile devices can utilize the data sharing and powerful computational resources In this paper, we investigate the cloud-based malware detection game, in used in mobile devices offload .We derive the Nash equilibrium (NE) and We designed malware detection scheme with Q-learning for a mobile device to derive the optimal offloading rate. The detection performance is improved with the Dyna architecture, in which a mobile device. We also design a post-decision state learning-based scheme that utilizes the known radio channel model Simulation results show that the proposed schemes improve the detection accuracy, reduce the detection delay, and increase the utility of a mobile device in the dynamic malware detection game, compared with the benchmark s

Richard Harang[26] states that to protect the intrusion detection system in this hidden markov model is used to examine the output of the model Kolmogorov test is conducted which produces best result

Vasileios Karyotis[27]: In order to provide importance to the network infrastructure from malware maintaining the communication networks (CCNs) from social coherency is importance .for that Markav random field method is used. MRF capture SIS malware propagation.SIS model is suitable for analysing malware propagation. Finally studying the cyber-physical systems can be studied using two-tier MRF

Chaitrali Amrutkar[28] stated that kayo mechanism is used to detect the mobile malicious webpage. We discover the malicious webpage which is missed by Google but this process is done using static malware analysis it does not produces better result so dynamic malware analysis technique is preferred.

Luca Caviglione,[29] proposed that this paper aim to spot the malware that changes rapidly by using artificial intelligence such as neural network and decision tree but experimentally it provides best result for detecting hidden data this process is mainly focused on colliding application but it communicates outside the sand box. Future work is based on detecting runtime scenario

RongWang[30] Malicious webpage are the serious threat to the society the static analysis is used to detect the malicious webpage it checks on unknown webpage but in this we combine both the static and dynamic analysis to determine whether hybrid combine with static shell code technique is used

Longfei Wu[31] mobile computing platform faces a serious threat to society due to phishing attack In order to overcome the phishing attack MobiFish technique is used in this we identified heuristic based anti-phishing scheme is that rely on html code but mobifish resolved this issue using OCR .It is implemented on real time in Google nexus.

Ambra Demontis[32]The number of threats have been in day to day life to overcome this machine learning technique is used drebin can be executed directly on mobile device the extracted malware train the classifier on the data labelled even though drebin provide best result but might exhibits vulnerability it provide way for the attacker limitation here is obfusion technique is used to detect Sec –SVM

Mahmudur Rahman[33] stated that Android is popular not only for its app but also for its malware proliferation. Previous work focus on android app executable and permission analysis .Fair play is a novel method to detect search rank fraud in Google play it provide a better result compare to previous one

Ethan M. Rudd[34] stated that government health care have become more digitized even though technology has been developed malware attack also developed at the same time in this survey paper malware detection method are categorized .while the machine learning algorithm offer anonymous solution this adaptive open world framework is used to recognize and relese it

Ying-Dar Lin, Chia-Yin Lee, Yu-Sung Tsai, [35] National Centre for High-Performance Computing proposed that malware is the only threat to society that spread directly or in directly through URL passive and active honey spot method are used here. honey spot give the same answer when in this we are able to identify the both the honey spot .honey spot execute malware in both strong and weak activities. But passive system is only in strong activity.

Sheng Wen, Student, Wei Zhou, Jun Zhang, Yang Xiang, Senior Wanlei Zhou, Senior, Weijia Jia[36], stated that reinfection and restart are two major malware characteristics .in this paper a novel difference equation is used and concept of virtual infected server is also used but in this paper is SII model is used to track the virtual infected user but it made independent assumption of user in the network and email is the main problem.

Michael Backes[37] states that leakage of the confidential data is the major threat to society which is faced by organisation but now it is faced by society in order to overcome this LIME protocol, is used in order to communicate between the two entity a novel algorithm is used in this every data is converted to image but it produces storage overhead

Abdulrahman alhothaily[38] states that in order to provide security to online banking authentication scheme using personal device is proposed it pays the way for user centric access and reduce the risk

III. PROPOSED WORK

To detect vulnerable and unsafe dynamic component the first automated technique was used in this paper, two phases analysis done here: 1) To collect runtime information on component loading we apply dynamic binary instrumentation (online phase); and 2) For detecting vulnerable component, we analyse the collected information (offline phase). In this we detect vulnerable and unsafe DLL loadings in popular Microsoft Windows software. Our results show that unsafe DLL loading can lead to serious security threats.

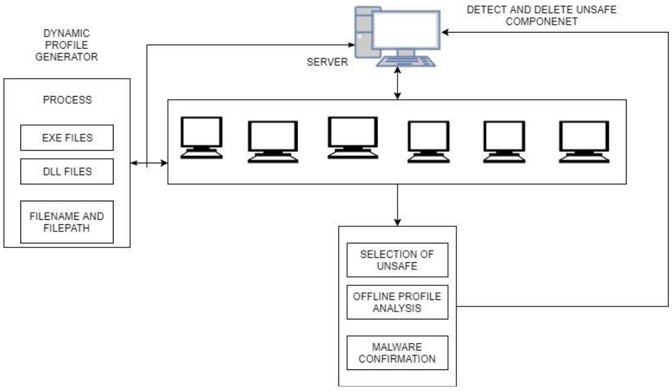

Fig.1 Proposed Architecture

Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar:

A. *Unsafe component loading*

Dynamic loading of components, is done here. Which follow the functionalities of shared libraries at runtime? Three phases of operation is done here: resolution, loading, and usage. Specifically, an application resolves the needed target components, loads them, and utilizes the desired functions provided by them. By dynamic loading the component interpretation can be done which is provided by runtime environments. . Full path and filename is the two types of target component specifications. For full path specification, operating systems resolve a target component based on the provided full path of the target component from the provided file name and a dynamically determined sequence of search directories. An OS iterates through the directories until it finds a file with the specified file name, which is the resolved component.

B. *Chained componenet loading*

In dynamic loading, the full path of the target component is determined by the resolution process and the component is incorporated into the host software if it is not already loaded

C. *Unsafe componenet resolution*

Dynamic loading is a critical step in software execution, but it has some security implication. Particularly loaded target component determined by the specified file name. This can lead to the loading of unintended components, it allow arbitrary code execution. Unsafe component resolution are classified into two types, resolution failure and unsafe resolution, and illustrated their conditions in Table 1.

IV. SUGGESTIONS FOR IMPLEMENTATION

In order to perform the malware analysis the researcher can perform operation using test bed, VMware virtual environment and Cuckoo Sandbox. Here this study shows clear that to use the machine learning algorithms to any problem, it's essential to represent the information in some type. For this purpose, Cuckoo Sandbox can be used. The reports generated by the sandbox, describing the activity information of every sample, were Pre-processed, and malware options were extracted from there. However, it's vital to know the practicality of the sandbox and also the structure of the reports initial. Cuckoo Sandbox is that the ASCII text file malware analysis tool that permits obtaining the elaborated activity report of any file or URL in an exceedingly matter of seconds. in keeping with Cuckoo Foundation currently, supported file formats include: Generic windows executable, DLL Files, PDF Documents, MS-Office documents, URLS and HTML Files, PHP Scripts, CPL Files, VB Scripts, ZIP Files, Java JAR, Python files and almost anything

V. CONCLUSION

Malware analysis is important in cyber security .many attacks have been generated .Increasing growth of malware aimed the researchers to develop a new technique to overcome the attacks, so this reason to choose machine learning.

REFERENCES

[1] Zhao, Z., Wang, J. and Bai, J., 2014. Malware detection method based on the control-flow construct feature of software. *IET Information Security*, *8*(1), pp.18-24.

[2] Bai, H., Hu, C.Z., Jing, X.C., Li, N. and Wang, X.Y., 2013. Approach for malware identification using dynamic behaviour and outcome triggering. *IET Information Security*, *8*(2), pp.140-151

[3] Das, S., Liu, Y., Zhang, W. and Chandramohan, M., 2016. Semantics-based online malware detection: towards efficient real-time protection against malware. *IEEE transactions on information forensics and security*, *11*(2), pp.289-302

[4] Zhang, F., Chan, P.P., Biggio, B., Yeung, D.S. and Roli, F., 2016. Adversarial feature selection against evasion attacks. *IEEE transactions on cybernetics*, *46*(3), pp.766-777.

[5] Zhao, G.U.O.D.O.N.G., Xu, K., Xu, L. and Wu, B., 2015. Detecting APT malware infections based on malicious DNS and traffic analysis. *IEEE Access*, *3*, pp.1132-1142.

[6] Aharoni, E., Peleg, R., Regev, S. and Salman, T., 2016. Identifying malicious activities from system execution traces. *IBM Journal of Research and Development*, *60*(4), pp.5-1.

[7] Hu, X., Jang, J., Wang, T., Ashraf, Z., Stoecklin, M.P. and Kirat, D., 2016. Scalable malware classification with multifaceted content features and threat intelligence. *IBM Journal of Research and Development*, *60*(4), pp.6-1.

[8] Gui, X., Liu, J., Chi, M., Li, C. and Lei, Z., 2016. Analysis of malware application based on massive network traffic. *China Communications*, *13*(8), pp.209-221.

[9] Lee, S., Kim, S., Lee, S., Yoon, H., Lee, D., Choi, J. and Lee, J.R., 2016. LARGen: automatic signature generation for Malwares using latent Dirichlet allocation. *IEEE Transactions on Dependable and Secure Computing*.

[10] Mazurczyk, W. and Caviglione, L., 2015. Information hiding as a challenge for malware detection. *arXiv preprint arXiv:1504.04867*.

[11] Fu, C., Liu, X.Y., Yang, J., Yang, L.T., Yu, S. and Zhu, T., 2017. Wormhole: The Hidden Virus Propagation Power of a Search Engine in Social Networks. *IEEE Transactions on Dependable and Secure Computing*.

[12] Carlin, D., Cowan, A., O'Kane, P. and Sezer, S., 2017. The Effects of Traditional Anti-Virus Labels on Malware Detection Using Dynamic Runtime Opcodes. *IEEE Access*, *5*, pp.17742-17752

[13] Tischer, M., Durumeric, Z., Bursztein, E. and Bailey, M., 2017. The Danger of USB Drives. *IEEE Security & Privacy*, *15*(2), pp.62-69

[14] Nadji, Y., Perdisci, R. and Antonakakis, M., 2017. Still beheading hydras: Botnet takedowns then and now. *IEEE Transactions on Dependable and Secure Computing*, *14*(5), pp.535-549.

[15] Courtney, M., 2017. States of cyber-warfare. *Engineering & Technology*, *12*(3), pp.22-25.

[16] Shen, Z., Hsu, C.W. and Shieh, S.W., 2017. Security Semantics Modeling with Progressive Distillation. *IEEE Transactions on Mobile Computing*.

[17] Islam, N., Das, S. and Chen, Y., 2017. On-Device Mobile Phone Security Exploits Machine Learning. *IEEE Pervasive Computing*, *16*(2), pp.92-96.

[18] Sun, M., Li, X., Lui, J.C., Ma, R.T. and Liang, Z., 2017. Monet: a user-oriented behavior-based malware variants detection system for android. *IEEE Transactions on Information Forensics and Security*, *12*(5), pp.1103-1112.

[19] Liu, B., Zhou, W., Gao, L., Zhou, H., Luan, T.H. and Wen, S., 2016. Malware Propagations in Wireless Ad Hoc Networks. *IEEE Transactions on Dependable and Secure Computing*.

[20] Mangialardo, R.J. and Duarte, J.C., 2015. Integrating Static and Dynamic Malware Analysis Using Machine Learning. *IEEE Latin America Transactions*, *13*(9), pp.3080-3087

[21] Saracino, A., Sgandurra, D., Dini, G. and Martinelli, F., 2016. Madam: Effective and efficient behavior-based android malware detection and prevention. *IEEE Transactions on Dependable and Secure Computing*.

[22] Abawajy, J. and Kelarev, A., 2017. Iterative classifier fusion system for the detection of Android malware. *IEEE Transactions on Big Data*.

[23] Ho, J.W. and Wright, M., 2017. Distributed Detection of Sensor Worms Using Sequential Analysis and Remote Software Attestations. *IEEE Access*, *5*, pp.680-695.

[24] Tian, K., Yao, D.D., Ryder, B.G., Tan, G. and Peng, G., 2017. Detection of Repackaged Android Malware with Code-Heterogeneity Features. *IEEE Transactions on Dependable and Secure Computing*.

[25] Xiao, L., Li, Y., Huang, X. and Du, X., 2017. Cloud-based Malware Detection Game for Mobile Devices with Offloading. *IEEE Transactions on Mobile Computing*.

[26] Harang, R. and Kott, A., 2017. Burstiness of Intrusion Detection Process: Empirical Evidence and a Modeling Approach. *IEEE Transactions on Information Forensics and Security*.

[27] Karyotis, V., 2017. A Markov Random Field Framework for Modeling Malware Propagation in Complex Communications Networks. *IEEE Transactions on Dependable and Secure Computing*

[28] Amrutkar, C., Kim, Y.S. and Traynor, P., 2017. Detecting mobile malicious webpages in real time. *IEEE Transactions on Mobile Computing*, *16*(8), pp.2184-2197.

[29] Caviglione, L., Gaggero, M., Lalande, J.F., Mazurczyk, W. and Urbański, M., 2016. Seeing the unseen: revealing mobile malware hidden communications via energy consumption and artificial intelligence. *IEEE Transactions on Information Forensics and Security*, *11*(4), pp.799-810.

[30] Wang, R., Zhu, Y., Tan, J. and Zhou, B., 2017. Detection of malicious web pages based on hybrid analysis. *Journal of Information Security and Applications*, *35*, pp.68-74.

[31] Wu, L., Du, X. and Wu, J., 2016. Effective defense schemes for phishing attacks on mobile computing platforms. *IEEE Transactions on Vehicular Technology*, *65*(8), pp.6678-6691.

[32] Demontis, A., Melis, M., Biggio, B., Maiorca, D., Arp, D., Rieck, K., Corona, I., Giacinto, G. and Roli, F., 2017. Yes, Machine Learning Can Be More Secure! A Case Study on Android Malware Detection. *IEEE Transactions on Dependable and Secure Computing*.

[33] Rahman, M., Rahman, M., Carbunar, B. and Chau, D.H., 2017. Search Rank Fraud and Malware Detection in Google Play. *IEEE Transactions on Knowledge and Data Engineering*, *29*(6), pp.1329-1342.

[34] Rudd, E., Rozsa, A., Gunther, M. and Boult, T., 2017. A survey of stealth malware: Attacks, mitigation measures, and steps toward autonomous open world solutions. *IEEE Communications Surveys & Tutorials*.

[35] Lin, Y.D., Lee, C.Y., Wu, Y.S., Ho, P.H., Wang, F.Y. and Tsai, Y.L., 2014. Active versus passive malware collection. *Computer*, *47*(4), pp.59-65.

[36] Wen, S., Zhou, W., Zhang, J., Xiang, Y., Zhou, W., Jia, W. and Zou, C.C., 2014. Modeling and analysis on the propagation dynamics of modern email malware. *IEEE transactions on dependable and secure computing*, *11*(4), pp.361-374.

[37] Backes, M., Grimm, N. and Kate, A., 2016. Data Lineage in Malicious Environments. *IEEE Transactions on Dependable and Secure Computing*, *13*(2), pp.178-191.

[38] Alhothaily, A., Hu, C., Alrawais, A., Song, T., Cheng, X. and Chen, D., 2017. A Secure and Practical Authentication Scheme Using Personal Devices. *IEEE Access*, *5*, pp.11677-11687.