Malware Detection: Issues and Challenges

Muchammad Naseer¹; Jack Febrin Rusdi¹; Nuruddeen Musa Shanono²; Sazilah Salam¹; Zulkiflee Bin Muslim³; Nur Azman Abu³; Iwan Abadi⁴.

¹ Informatics, Sekolah Tinggi Teknologi Bandung, Bandung, Indonesia.
² Kano University of Science and Technology, Wudil, Kano, Nigeria.
³ Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia.
⁴ Informatics Engineering, Universitas Langlangbuana, Bandung, Indonesia.

Corresponding author: inijack@gmail.com; jack@sttbandung.ac.id

Abstract. Malware is a severe threat that makes computer security more vulnerable. Many studies have been conducted to improve the capability of detection techniques. However, there is a lack of analysis of the current trend of IDS. This paper is about extracting and analyzing the latest detection techniques which had been conducted by various studies. This paper will also emphasize the current challenges of malware deployment from recent studies. Finally, the similarities and differences between the detection techniques will be exposed, and the issues and problems related to detection techniques will highlight as well. In the future, this paper outcome can be used to highlight the current topic addressed in malware research.

1. Introduction

In today's world of computing, Malware is a grave threat that makes computer security more vulnerable [1]. Generally, various malware activities firmly bound by the opening of information and networks that exist today [2], [3], including network management and settings [4], [5]. The act of malicious activity continues to grow exponentially, and it is getting very sophisticated [6], [7]. It exposes the computer system to the possibility of being attacked or harmed, employing Internet or data communication [8].

The internet and data communication, both through existing technology and various possibilities for future development [5], provide opportunities for the growth of this malware.

The term Malware comes from merging the two words “Malicious” and “Software.” A Malware is a piece of program that installed on a system without the knowledge of the owner to steal sensitive information, accesses private data, and also harms the system by altering the functionality of some legitimate applications which slows down the system [9]–[11].

Malware needs to be identified and removed from the infected system to avoid the leakage of sensitive data or any other malicious activity in the computer. Malware classified into viruses, worms, Trojans, Spywares, Adwares, Rootkits [12], [13].

In the next section, the remaining part of the paper divided into different sections. Section 2 highlights the malware overview. Section 2.1 presents the malware detection techniques classification. Sections 2.2 discussed the malware detection analysis. Section 3 discusses the issues and limitations. Lastly, the study concluded in Section 4.
2. Malware overview

Malware (short for malicious software), is usually considered as software that aims to disrupt regular activities of a computerized system by gathering sensitive information or making unauthorized access to computer systems and mainly irritate clients [13].

Malware is divided into virus, worm, and trojan [13], [14].

A computer virus is a code that replicates when inserting into other programs. A necessary caution is that in order for viruses to function, a virus needs an existing host program for it to cause harm [9]. Unlike a virus, computer worm replicates itself by executing its code independent of any other program [11]. In general, viruses attempt to spread through programs/files on a single computer system [15]. While worms spread through network connections to infect as many computer systems connected to the network as possible [9], and strengthened by the high reliance on the use of technology for humans today [16]. A Trojan horse is a malware embedded by its designer in an application or system. The application or system would appear to perform some useful function but is performing some unauthorized action like capturing the user’s information through keystrokes and sending it to a malicious host. [9], [17]

Even though there are lots of security solutions available in the current market like antivirus, SSL certificate encryption, firewall protection, these security solutions only provide temporary protection due to their defensive mode. Hence, the new defensive solution must be updated consistently to ensure it continuously protects the information against malicious activities or software.

Recently, through universities as higher education institutions and research institutions [18], [19], various studies have proposed new techniques and approaches to solve these problems, such as limited storage, which resulted in the existence of a large amount of literature in the field of study. In this paper, we conducted a review study of the existing literature in the field of study to highlights the existing research issues and challenges. The outcome of our work shows that they need to reduce signature repository and increase accuracy score.

Table 1: Malware Detection Techniques Classification

| Detection techniques | Definition | Benefit | Limitation |
|----------------------|------------|---------|------------|
| Signature-based      | It is the most generally utilized antivirus method. A signature is a succession of bytes that can be utilized to distinguish malware. [8], [12], [20], [21] | - Straight forward and relatively fast - Successful against most regular sorts of malware. | - Requires a forward mark database as malware not present in the database will not be recognized. - Relatively basic obscurity system can be utilized to dodge this method. |
| Behavior-based       | It centers around the activities performed By the malware amid execution [6], [12], [22] | Systematical conduct investigation of the suspected malware | - Both benign and malware examined amid the preparation stage. -Classification of them will be only during the execution phase. |
| Statistical based    | Properties derived from program features as in Hidden display are utilized to arrange transformative/metamorphic malware [23] | Has served a benchmark in an assortment of different investigations | Visible only when utilizing HMMs as the basis for the malware detection schemes |
| Heuristic Techniques | Primarily utilize machine learning and data mining strategies to recognize the conduct of the running project. [12], [13] | -Distinguished polymorphic and obscure malware. -Fewer false positive than different scanners. | - Expansive arrangement of produced rules for building classifier. |
| Anomaly-based        | Typically happens in two stages, a preparation (learning) stage and an identification (monitoring) stage. [7], [10], [24]–[26] | -Increasing the rule set helps in less false positive alarms. -A novel attack for which a signature does not exist can be recognized. | - Ability to be tricked by an effectively conveyed assault -A high false-positive ratio |
2.1. Malware Detection Techniques Classification

2.2. Classification of Malware detection techniques seen from several aspects, such as those based on Signature-based, Behavior-based, Statistical-based, Heuristic techniques, and anomaly-based. This classification technique, as shown in Table 1.

Table 2: Malware Detection Analysis

| Author            | Analysis                                                                 | Design                                                                 | Implementation                                                                 | Testing                                                                 |
|-------------------|---------------------------------------------------------------------------|------------------------------------------------------------------------|------------------------------------------------------------------------------|------------------------------------------------------------------------|
| Huda et al. (2018) | Information sent to the sandbox to produce a behavioral log file.         | Data will be tested to produce the extracted run time feature, file training, and test data. | Framework for feature selection, estimate model parameters for wrapper, most significant run time behavior. | Wrapper detection engine will verify whether result is harmful or not |
| Narudin et al. (2016) | The first stage is data collection, which captures the network traffic of normal and malicious applications and transmits it to the next phase. | Feature selection and extraction, the TCP packets are filtered; then, features selected from among various network features extracted, labeled, and stored in a database to applied in the next phase. | The machine learning classifier entails the final phase, whereby the information in the database trains the machine learning classifier to produce a detection model. | |
| Noor et al. (2018) | A malware sample executed in a sandbox, which based on Cuckoo malware analysis engine. | AEMS employs a built-in-the box monitoring mechanism in the form of a dynamic link library deployed as a kernel driver. | The results are generated in the form of an execution profile, describing the malware behavior categorized in the network activity, file system activity, and system calls. | |
| Talha et al. (2015) | APK Auditor client offers an analysis request, showing whether the application can be trusted or not. | Applications are stored in the APK Auditor signature database server together with the results of the analysis. | The APK Auditor central server manages the analysis process and works as a link between the signature database and the Android client while analyzing requested applications. | |
| Ambusaidi et al. (2016) | Data collection is where sequences of network packets are collected | Data preprocessing, is where training and test data are preprocessed and essential features that can distinguish one class from the others are selected | Classifier training is where the model for classification is trained using LS-SVM | Attack recognition is where the trained classifier is used to detect intrusions on the test data. |
| Ali Mirza et al. (2018) | Prepares file by eliminating any apparent obfuscation, it is then thoroughly analyzed and all the possible features are | Obtfuscated parts from eliminated, and the features arranged in a format that is understandable by the classification module. | Analysis result of module output stored in the analysis repository. | |
| Tong & Yan (2017) | Data gathered about runtime system calls of a set of known malware and benign apps using a dynamic method in order to traverse most app features | Malicious and a regular pattern set built by extracting the patterns from the collected data | Different patterns between malicious and benign apps inserted into the malicious pattern set, and different patterns between normal and malicious apps inserted into the familiar pattern set. | The runtime system calling data collected about both individual system calls, and sequential system calls with different depth, and then the target patterns are extracted. |

2.3. Malware Detection Analysis

2.3.1. Malware Detection Construction. Analysis of the detection of construction of Malware was carried out by several researchers, as shown in Table 2. The criteria for detecting this construction include review based on Design, Implementation, and Testing.
2.3.2. Malware Detection Characteristics

2.3.3. Malware has several characteristics. Related to the characteristics of malware, some of which are seen based on novelty, advantages, and disadvantages based on previous research, as shown in Table 3.

| Author | Novelty | Advantages | Disadvantages |
|--------|---------|------------|---------------|
| Huda et al. (2018) [27] | A hybrid-multi filter-wrapper based framework was proposed that overcomes the limitations of current detection systems. - It integrates the knowledge (from the intrinsic Characteristics of run-time behavior) obtained by more than one filters into the wrapper. | Improves the detection accuracies by taking advantages of the filter and wrapper. - The proposed framework finds the most significant run-time characteristics of malware. | All hybrids in the framework reduce the computational complexity from an exponential to a polynomial type as a function of the cardinality of the run-time features sets of malware. |
| Narudin et al. (2016) [28] | This study presented an evaluation using machine learning classifiers to detect mobile malware effectively by selecting the appropriate network features for inspection by the classifiers, as well as to determine the ideal classifier based on true-positive rate (TPR) values. | Proves the effectiveness and efficiency of machine learning in a real mobile malware operational environment. - proves that machine learning classifiers can detect the latest malware. | Evaluation is limited to a certain amount of malware, and few of the approaches consider feature selection in the classification process to increase the result accuracy. |
| Noor et al. (2018) [29] | Analysis Evasion Malware Sandbox - AEMS system, which possesses the capability of detecting the presence of an analysis evasion technique within the malware and can force it to exhibit its correct functionality inside the sandbox. | Effective in detecting the variations in malware behavior within the sandbox and the proof of concept countermeasures, implemented by AEMS are effective against a large proportion of common malware. - A novel technique for detection of malware evasive behavior is Presented. | AEMS faces constraints in offsetting the effects of evasion based on timing differences and through the identification of the parent process. - it less scalable for handling a large number of malware samples which generated every day |
| Talha et al. (2015) [30] | Provides a new approach to assessing potential the maliciousness of Android applications by calculating a statistical score through the requested permissions - Application analysis is wholly carried out on a central server, and the results retrieved by a web service | APK Auditor is a learning-based, extensible, and lightweight system that provides a new approach for Android malware detection. - Helps digital investigators, and likely Android users to check whether or not applications are malicious. | Because it is a learning-based mechanism, each application analysis affects the malware detection process positively moreover updates the signature database. - The number of FP and FN detections is high. |
| Ambusaidi et al. (2016) [31] | A new filter-based feature selection method, in which theoretical analysis of mutual information is introduced to evaluate the dependence between features and output classes. - An (IDS), named Least Square Support Vector Machine based IDS (LSSVM-IDS), is built. | The proposed detection system has achieved promising performance in detecting intrusions over computer networks. -A flexible method for the problem of feature selection, FMIFS, is developed. - The proposed feature selection algorithm is computationally efficient when it applied to the LSSVM-IDS. | The proposed feature selection algorithms can only rank features in terms of their relevance, but they cannot reveal the best number of features that are needed to train a classifier. -The proposed feature selection algorithm could be further enhanced by optimizing the search strategy. |
| Ali Mirza et al. (2018) [32] | One of the contributions of this paper is energy efficiency, which is one of the weakest areas of many antivirus. -The classification methodology proposed in this research prove the initial hypothesis of enhanced accuracy in malware identification | An approach that applied on multiple security threats and can identify not just known, but it can also predict unknown threats. - Capable of managing a large number of requests coming from multiple individual clients and enterprise networks. | Unable to compare the systems overall performance results with any previously available study. - They did not test the cloud-based architecture against a large number of clients or a big network. |
The above seven latest malware detection technique was selected and analyzed. We then extracted the relevant information and compared, and then finally a conclusion made.

3. Issues and Challenges
In this section, we present some of the issues and challenges identified in the field of study which is, for example;

| Stages     | Issues                                                                 |
|------------|------------------------------------------------------------------------|
| Data Collection | - Few of the approaches consider feature selection in the classification process to increase result accuracy. |
| Analysis   | - Some of the evaluations are limited to a certain amount of malware like anomaly-based approach. |
|            | - Some approaches could not reveal the best number of features that are needed to train a classifier. |
| Response   | - Scalability in handling a large number of malware samples. |
|            | - The high number of false positive and false negative. |
|            | - Need to gather new malware and make it benign continually. |
|            | - Limited computing and storage resources. |

4. Conclusion
Malware has rapidly become a significant security threat for the computing community, which becomes one of the reasons for most of the current security problems on the Internet. Although a considerable amount of research effort has gone into malware detection, however, malicious code remains a vital threat on the Internet today. Of recent, various Malware detection techniques and approaches have been proposed to tackle these problems. Unfortunately, these techniques and approaches have some shortcomings that deter them from eliminating the problem.

This paper extracts the shortcomings of the latest detection techniques for further analysis. The outcome of our work shows that there is a need to reduce signature repository fit in lightweight devices such as IoT sensors. Our future work is to deal with the problem.

Acknowledgment
This research conducted by the Pervasive Computing & Educational Technology Research Group, C-ACT, Universiti Teknikal Malaysia Melaka (UTeM). Sekolah Tinggi Teknologi Bandung which has provided research materials.

References
[1] A. Qamar, A. Karim, and V. Chang, “Mobile malware attacks: Review, taxonomy & future directions,” Future Generation Computer Systems, vol. 97, pp. 887–909, Aug. (2019), doi: 10.1016/J.FUTURE.2019.03.007.
[2] J. Febrian, “Menjelajah Dunia dengan Google,” Penerbit Informatika, (2008).
[3] J. Febrian, “Google & Yahoo Secrets!,” Penerbit Informatika, (2007).
[4] M. R. K. Ariffin, M. A. Asbullah, and N. A. Abu, “Security Features of an Asymmetric Cryptosystem based on the Diophantine Equation Hard Problem,” Mar. (2011).
[5] J. F. Rusdi, S. Salam, N. A. Abu, S. Sahib, M. Naseer, and A. A. Abdullah, “Drone Tracking Modelling Ontology for Tourist Behavior,” Journal of Physics: Conference Series, vol. 1201, no. 1, p. 012032, May (2019), doi: 10.1088/1742-6596/1201/1/012032.
[6] A. Souri and R. Hosseini, “A state-of-the-art survey of malware detection approaches using data mining techniques,” Human-centric Computing and Information Sciences, vol. 8, no. 1, p. 3, Dec. (2018), doi: 10.1186/s13673-018-0125-x.
[7] Jabez J and B. Muthukumar, “Intrusion Detection System (IDS): Anomaly Detection using Outlier Detection Approach,” Procedia - Procedia Computer Science, vol. 48, pp. 338–346, (2015), doi: 10.1016/j.procs.2015.04.191.
[8] D. Gavrilit, M. Cimpoesu, D. Anton, and L. Ciortuz, “Malware detection using machine learning,” in 2009
International Multiconference on Computer Science and Information Technology, 2009, pp. 735–741.

[9] N. Idika and A. P. Mathur, “A Survey of Malware Detection Techniques,” SERC Technical Reports, no. October, p. 48, (2007).

[10] ONT209, “Malware Detection Techniques Description | MalwareTips Community,” Malwaretips, 2013. [Online]. Available: https://malwaretips.com/threads/malware-detection-techniques-description.14028/. [Accessed: 31-Aug-2019].

[11] M. A. Jerlin and K. Marimuthu, “A New Malware Detection System Using Machine Learning Techniques for API Call Sequences,” Journal of Applied Security Research, vol. 13, no. 1, pp. 45–62, Jan. (2018), doi: 10.1080/19361610.2018.1387734.

[12] S. Alqurashi and O. Batarfi, “A Comparison of Malware Detection Techniques Based on Hid-den Markov Model,” Journal of Information Security, vol. 7, pp. 215–223, (2016), doi: 10.4236/jis.2016.73017.

[13] Z. Bazrafshan, H. Hashemi, S. M. H. Fard, and A. Hamzeh, “A survey on heuristic malware detection techniques,” in The 5th Conference on Information and Knowledge Technology, 2013, pp. 113–120.

[14] M. Christodorescu, S. Jha, S. A. Seshia, D. Song, and R. E. Bryant, “Semantics-Aware Malware Detection,” in 2005 IEEE Symposium on Security and Privacy (S&amp;P’05), pp. 32–46.

[15] M. I. Abdullah Almarshad, M. M. Z. E Mohammed, and A.-S. Khan Pathan, “Detecting Zero-day Polyorphic Worms with Jaccard Similarity Algorithm,” (2016).

[16] J. F. Rusti, S. Salam, N. A. Abu, B. Sunaroyo, R. Taufiq, L. S. Muchlis, T. Septiana, K. Hamdi, Arianto, B. Ilman, Desfitriady, F. R. Kodong, and A. V. Vitaningsih, “Dataset Smartphone Usage of International Tourist Behavior,” Data in Brief, p. 104610, Oct. (2019), doi: 10.1016/j.dib.2019.104610.

[17] B. Amro, “Malware Detection Techniques for Mobile Devices,” International Journal of Mobile Network Communications & Telematics, vol. 7, no. 4/5, pp. 01–10, Dec. (2017), doi: 10.5121/ijmncett.2017.7601.

[18] J. Febrian, “Buku Saku Tentang Pendidikan Tinggi di Indonesia,” Penerbit Informatika, (2000).

[19] J. F. Rusti, S. Salam, N. A. Abu, T. G. Baktina, R. G. Hadiningrat, B. Sunaroyo, A. Rusmartiana, W. Nashiuhuddin, P. Fannya, F. Laurenty, N. Shanono, and R. Hardi, “ICT Research in Indonesia,” SciTech Framework, vol. 1, pp. 1–23, (2019).

[20] P. Pongle and G. Chavan, “A survey: Attacks on RPL and 6LoWPAN in IoT,” in 2015 International Conference on Pervasive Computing (ICPC), 2015, pp. 1–6.

[21] S. G. Kene and D. P. Theng, “A review on intrusion detection techniques for cloud computing and security challenges,” in 2015 2nd International Conference on Electronics and Communication Systems (ICECS), 2015, pp. 227–232.

[22] P. D. Sawle and A. B. Gadicha, “Analysis of Malware Detection Techniques in Android,” (2014).

[23] G. A. N. Mohamed and N. B. Ithnin, “Survey on Representation Techniques for Malware Detection System,” American Journal of Applied Sciences, vol. 14, no. 11, pp. 1049–1069, Nov. (2017), doi: 10.3844/ajassp.2017.1049.1069.

[24] V. Jyothsna and V. V. R. Prasad, “A Review of Anomaly based IntrusionDetection Systems,” International Journal of Computer Applications, vol. 28, no. 7, (2011).

[25] A. Sari, “A Review of Anomaly Detection Systems in Cloud Networks and Survey of Cloud Security Measures in Cloud Storage Applications,” Journal of Information Security, vol. 06, no. 02, pp. 142–154, Mar. (2015), doi: 10.4236/jis.2015.62015.

[26] N. M. Zamry, A. Zainal, and M. A. Rassam, “Unsupervised Anomaly Detection for Unlabelled Wireless Sensor Networks Data,” (2018).

[27] S. Huda, R. Islam, J. Abawajy, J. Yearwood, M. M. Hassan, and G. Fortino, “A hybrid-multi filter-wrapper framework to identify run-time behaviour for fast malware detection,” Future Generation Computer Systems, vol. 83, pp. 193–207, Jun. (2018), doi: 10.1016/J.FUTURE.2017.12.037.

[28] F. A. Narudin, A. Feizollah, N. B. Anuar, and A. Gani, “Evaluation of machine learning classifiers for mobile malware detection,” Soft Computing, vol. 20, no. 1, pp. 343–357, Jan. (2016), doi: 10.1007/s00500-014-1511-6.

[29] M. Noor, H. Abbas, and W. Bin Shahid, “Countering cyber threats for industrial applications: An automated approach for malware evasion detection and analysis,” Journal of Network and Computer Applications, vol. 103, pp. 249–261, Feb. (2018), doi: 10.1016/J.JNCA.2017.10.004.

[30] K. A. Talha, D. I. Alper, and C. Aydin, “APK Auditor: Permission-based Android malware detection system,” Digital Investigation, vol. 13, pp. 1–14, Jun. (2015), doi: 10.1016/J.DIIN.2015.01.001.

[31] M. A. Ambusaidi, X. He, P. Nanda, and Z. Tan, “Building an Intrusion Detection System Using a Filter-Based Feature Selection Algorithm,” IEEE Transactions on Computers, vol. 65, no. 10, pp. 2986–2998, Oct. (2016), doi: 10.1109/TC.2016.2519914.

[32] Q. K. Ali Mirza, I. Awang, and M. Younas, “CloudIntell: An intelligent malware detection system,” Future Generation Computer Systems, vol. 86, pp. 1042–1053, Sep. (2018), doi: 10.1016/J.FUTURE.2017.07.016.

[33] F. Tong and Z. Yan, “A hybrid approach of mobile malware detection in Android,” Journal of Parallel and Distributed Computing, vol. 103, pp. 22–31, May (2017), doi: 10.1016/J.JPDC.2016.10.012.