AN IMPROVED MALWARE DETECTION FRAMEWORK

AHMAD NAIM IRFAN BIN ASWAMI FADILLAH

UNIVERSITI TEKNOLOGI MALAYSIA

AN IMPROVED MALWARE DETECTION FRAMEWORK

AHMAD NAIM IRFAN BIN ASWAMI FADILLAH

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Master of Philosophy

Razak Faculty of Technology and Informatics
Universiti Teknologi Malaysia

OCTOBER 2020

DEDICATION

This thesis is specially dedicated to my beloved father, Dr Aswami Ariffin (Dr AA) who is my idol, mentor and source of inspiration. Also to my beloved mother, Anne who has provided guidance and strength in completing this thesis. Both of them have provided endless encouragement and support to make this endeavour possible.

To my siblings, Nadzirah, Nabilah, Nadia and Najwa who have been by my side during my highs and lows.

And lastly, praise be to the Almighty God, this thesis is finally completed. May this be the start of an adventure of a lifetime.

ACKNOWLEDGEMENT

Foremost, I would like to express my sincere gratitude to my supervisor, Ap. Ts. Dr. Mohd Naz'ri Bin Mahrin for the continuous support of my research, for his knowledge, patience, and guidance in helping me write this thesis. Also, Dr. Mohd Syahid Bin Mohd Anuar for his insights, time, and guidance in this research.

Lastly, I am also thankful to Universiti Teknologi Malaysia (UTM) especially the UTM staffs that have been directly or indirectly involved in this research and helped me throughout the preparation of the years of my study.

ABSTRACT

The detection of malware intrusion requires the identification of its signature. However, cyber security practitioners are having difficulty to manually detect signature-based malware due to the increasing number of malware. As a consequence, malware are only detected after an incident has occurred. By then it would have already incurred monetary loss, thus causing a huge impact on an organisation's brand and clients' trusts. This research aims to propose a solution for the problem highlighted by formulating an improved malware detection framework. The improved malware detection framework was formulated based on the malware detection solution components identified as malware analysis, malware detection, machine learning algorithm, cyber threat intelligence data and digital forensics principle (preservation). Then, the formulated framework was implemented and evaluated by performing a threat hunting experiment. The implementation of the formulated framework produced information that described the distribution of high severity malware which posed the most threat in the top three states based on the clustering algorithm used. The clustering algorithm used 3 as the value of K which had the best silhouette score based on Euclidean distance calculated that is 0.931766381586 and assisted in generating the YARA rules. The experiment result shows that the generated YARA rules from the clustering algorithm and data enrichment were able to detect Bladabindi, Conficker as well as Zbot by referring to the signature derived from the automated malware analysis. As a conclusion, the framework itself, steps, techniques and the process flow utilised in formulating the improved framework served as an effective malware detection solution. Hence, cyber security practitioners can apply the improved malware detection framework as a guideline to conduct threat hunting within their organisation.

ABSTRAK

Pengesanan pencerobohan perisian merbahaya memerlukan pengenalpastian *signature*. Walau bagaimanapun, pengamal keselamatan siber mengalami kesukaran untuk mengesan perisian merbahaya berdasarkan *signature* secara manual kerana jumlah perisian merbahaya yang semakin meningkat. Akibatnya, perisian merbahaya hanya dapat dikesan selepas berlakunya kejadian. Pada masa itu, ia akan mengalami kerugian kewangan, sehingga menimbulkan impak basar pada jenama organisasi dan kepercayaan pelanggan. Kajian ini bertujuan untuk mencadangkan penyelesaian bagi permasalahan yang dinyatakan dengan merumuskan kerangka pengesanan perisian merbahaya yang lebih baik. Kerangka pengesanan perisian merbahaya yang lebih baik dirumuskan berdasarkan komponen penyelesaian pengesanan perisian merbahaya yang dikenal pasti sebagai analisis perisian merbahaya, pengesanan perisian merbahaya, algotrima pengesanan mesin, data risikan ancaman siber dan prinsip forensik digital (pemeliharaan). Kemudian, rangka kerja yang dirumuskan dilaksanakan dan dinilai melalui eksperimen *threat hunting*. Pelaksanaan rangka kerja yang dirumuskan menghasilkan maklumat yang menghuraikan pengagihan perisian merbahaya berketerukkan tinggi yang menimbulkan ancaman paling besar dalam tiga negeri teratas berdasarkan algoritma pengklusteran yang digunakan. Algoritma pengklusteran menggunakan 3 sebagai nilai K yang mempunyai *Silhouette score* tertinggi berdasarkan *Euclidean distance* yang dikira iaitu 0.931766381586 dan membantu untuk menghasilan peraturan YARA. Keputusan eksperimen menunjukkan bahawa peraturan YARA yang dihasilkan dari algoritma pengklusteran dan pengayaan data dapat menggesan Bladabindi, Conficker dan juga Zbot dengan merujuk kepada *signature* yang diperoleh dari analisis data secara automatik. Sebagai kesimpulan, kerangka kerja tersebut, langkah, teknik dan aliran proses yang digunakan dalam merumuskan kerangka pengesanan perisian merbahaya dapat diguna pakai sebagai langkah penyelesaian pengesanan perisian merbahaya yang efektif. Oleh itu, pengamal keselamatan siber dapat menerapkan rangka kerja pengesanan perisian merbahaya yang ditingkatkan sebagai panduan bagi melakukan *threat hunting* dalam organisasi mereka.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

API	-	Application Programming Interface
APT	-	Advanced Persistent Threat
CSV	-	Comma-Separated Values
CTIP	-	Cyber Threat Intelligence Program
CVE	-	Common Vulnerabilities and Exposures
GMM	-	Gaussian Mixture Model
GB	-	GigaByte
GeoIP	-	Geolocation Internet Protocol
HDFS	-	Hadoop Distributed File System
ICT	-	Information and Communication Technology
IDS	-	Intrusion Detection System
IPS	-	Intrusion Prevention System
IOC	-	Indicator Of Compromise
IoT	-	Internet of Things
IP	-	Internet Protocol
JSON	-	Java Script Object Notation
MISP	-	Malware Information Sharing Platform
PC	-	Personal Computer
PE	-	Portable Executable
REST	-	Representational State Transfer
UI	-	User Interface
UML	-	Unsupervised Machine Learning
VM	-	Virtual Machine

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Overview

This chapter gives the outline of the research that includes the background explanation and problem encountered by cyber security practitioners in malware detection. In addition to that, the research scope is discussed with consideration of the challenges in the research. Apart from that, the proposed solution is briefly described in the research contribution and significance of the study sections. The overview for each chapter is provided at the end of this chapter.

1.2 Research Background

Technology assists humans to solve problems and perform difficult tasks that humans are incapable of, for example, when running an algorithm to solve a complex mathematical problem within seconds. The birth of the internet has made the world borderless that enables tasks to be carried out on mobile technology namely laptops and smartphones. The internet allows tasks to be performed online at any time and place, for example, performing online banking through mobile devices. However, technologies that exist for the benefit of humans are misused by cybercriminals that pose harm and threat to others through cybercrimes that involve malware (Deckert & Sarre, 2017). One example of cybercrime is hacking a victim's device to steal their personally identifiable information (PII) by attaching malware in the victim's email that has been sent to them (Gunjan et al., 2013). Although technology advancement has improved the quality of human life, it also causes cybercrimes to become more sophisticated through cyber-attacks, to create better malware that is more elusive than the previous version (Hopkins & Dehghantanha, 2016).

Cyber-attacks using malware are prevalent to automate the process of intrusion into the targeted organisation's IT systems, such as in the financial sector. Zero-days, vulnerabilities found within applications and operating systems without patching are often utilised by cyber criminals to create a new malware variant to avoid security tools deployed by cyber security practitioners. Malware is able to evade the security systems by changing its signature so that it is not included in the antivirus repository for reference. An example of this issue is polymorphic and metamorphic malware that are undetected by applying a packing technique to mask its original behaviour (Bat-Erdene et al., 2017). Malware authors currently develop features that are difficult to trace such as the existence of obfuscation technique (Raphel & Vinod, 2015) that complicate detection and reverse engineering. Obfuscation technique hides the malicious intent of the malware and masks the malware as a legitimate software (Martinelli et al., 2018). However, in the background, the malware performs illegal processes such as privilege escalation by encrypting malicious code to appear as legitimate. This shows the sophistication level of the techniques being deployed for malware intrusion.

Generally, the malware intrusion phases consist of reconnaissance, weaponisation, delivery, exploitation and installation to intrude and have total control over the victim's systems (Kiwia et al., 2018). The attack model customisable by changing the technique used in the malware intrusion phases depending on the scenario and its target (Bhatt et al., 2014). Therefore, as depicted in the McAfee December 2018 threats report, new malware samples jumped in quarter three of 2018 to approximately 63 million which saw a 53% increase (Boom et al., 2019). Malware consists of many variations and the current technology such as antivirus is incapable of detecting new malware intrusion as it relies on a predefined repository. Apart from that, solely depending on antivirus is not recommended as it consumes a lot of resources and fails to completely secure the network or system (Ali Mirza et al., 2018).

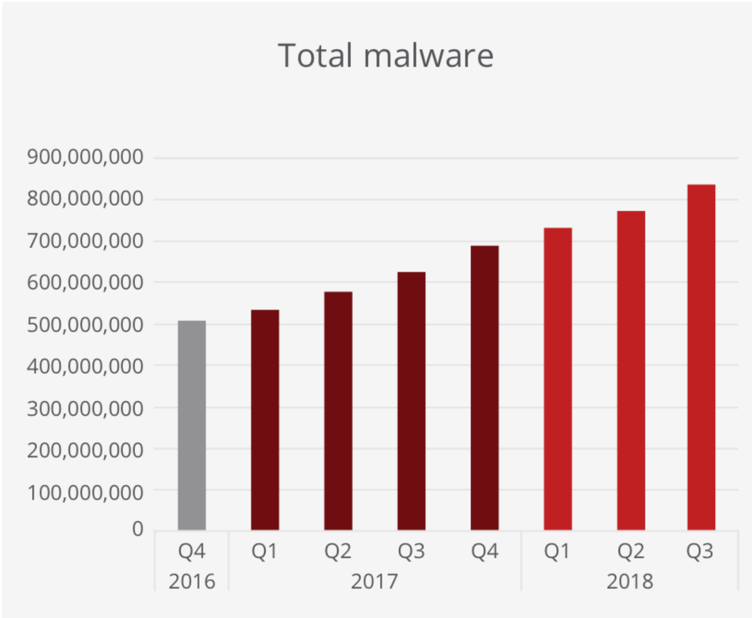


Figure 1.1 Total malware reported by McAfee Labs (2016-2018)

Figure 1.1 shows the total number of malware estimated by McAfee Lab in each quartile of the year from 2016 to 2018 based on the reports issued by McAfee Labs (Boom et al., 2019). Based on the McAfee December 2018 threats report, it is evident that the number of malware increases each year and that cyber security practitioners are dealing with millions of malware each year. This is because new malware is created faster than the analysis performed by malware analysts (Bulazel & Yener, 2017). The problem is not because of the deficiency in malware analysis techniques but rather the manual malware analysis that depends on human is time consuming. Conducting malware analysis manually is beyond the human's capability. Thus, security tools exist to help ease malware analysts' burden. Tools such as VirusTotal have the functionality to produce a report on the file scanned and detect malware e.g. based on the hash values of the file (VirusTotal, n.d.). Nevertheless, a malware detection solution cannot be too dependent on tools as malware evolves and tools may become obsolete.

Apart from using security tools to discover malware, malware detection is improved through the utilisation of machine learning algorithm during malware analysis (Mohaisen et al., 2015). The process of malware analysis is automated through the inclusion of machine learning algorithm to identify malware signatures. For example, an automated malware analysis is implemented using machine learning in the dynamic analysis approach for the discovery of out of the norm system call (Naval, Laxmi, Rajarajan, et al., 2015). The example shows that malware analysis on more than one malware at a time is possible using machine learning algorithms e.g. in determining the outlier of a behaviour to recognise malicious activity (Ajay Kumara & Jaidhar, 2017). In addition to that, machine learning algorithm and big data system are technologies used to counter cyber-attacks (Kozik, 2018) and improve the detection of malware (Incer et al., 2018). The data collected is improved through data enrichment and malware analysis using machine learning algorithm in the extraction of malware characteristics (Martinelli et al., 2017). The extracted malware characteristics or behaviour are information which is used in the generation of malware signature to detect malware. Apart from analysing malware, cyber threat intelligence is used to obtain information of malware signatures.

Cyber threat intelligence is utilisable as an external data for malware signature repository enrichment and evidently as the malware analysis main source of information. This is because cyber threat intelligence data contains attack details on present and arising threats (Abu, Selamat, Ariffin, et al., 2018). It is considered as inefficient if the malware analysis is performed on discovered malware. Resources are wasted to obtain the same malware analysis result e.g. malware report on the malware signature to deal against the same malware. Instead, malware analysis results shared by the cyber security community should be utilised. Static and dynamic analysis are the two main malware analysis techniques used to obtain the malware signature. However, malware analysis conducted may not be thorough and requires additional information such as malware Internet Protocol (IP), malicious Domain Name System (DNS) and other reputation data to generate malware signatures through enrichment. Therefore, the data regarding malware is not only obtainable from malware analysis but also from external cyber threat intelligence data that collects network and system data. An example of external cyber threat intelligence data is the data collected from the VirusTotal database that contains data such as malware IP, malicious DNS as well as other reputation data. The collection of cyber threat intelligence data provides the means to establish the understanding of cyber threat in an environment which is complemented by the deployment of machine learning algorithm to automate malware analysis for the detection of malware.

Cyber threat intelligence data contains vital malware information and preservation of the data is essential in ensuring the accessibility of the data (Okereke & Chukwunonso, 2018). Preservation is a principle in digital forensic to ensure that the original data is retrievable to prevent data loss and alteration to the original data (Luthfi & Prayudi, 2016). This is conducted to ensure the data integrity where the data is preserved before analysis is conducted so that the analysis is carried out on the copy of the original data i.e. changes are not made to the original data. Therefore, preservation would be ideal for analysis that involves handling crucial data.

Malware detection as early as possible is crucial as most malware incidents are detected by the organisation only after experiencing visible consequences such as unauthorised money transfer and down of services as stated in the Kaspersky Incident Response Analytics Report 2018 (Kaspersky, 2018). By then, it is already too late as the impact from the malware incident is huge which affects the organisation brand as well as the organisation client's trust apart from the monetary loss (Pandey et al., 2020).

1.3 Research Problem

Since the number of malware increases yearly, cyber security practitioners are having difficulty to manually perform signature-based malware detection. y then it would have already incurred monetary loss, thus causing a huge impact on an organisation's brand and clients' trusts. To address this problem, a malware detection framework that performs malware analysis through the integration of machine learning algorithm, cyber threat intelligence and digital forensics principle (preservation) as a malware detection reference for researchers and cyber security practitioners.

1.4 Research Questions

The research questions of the study are as follows:

i. What are malware detection solution components?
ii. How to formulate an improved malware detection framework based on the identified malware detection solution components?

iii. How to evaluate the improved malware detection framework formulated by a threat hunting experiment?

1.5 Research Objectives

The research objectives that guides the study are as follows:

i. To identify malware detection solution components.
ii. To formulate an improved malware detection framework based on the identified malware detection solution components.
iii. To evaluate the improved malware detection framework by performing a threat hunting experiment.

1.6 Research Scope

There are a few challenges encountered in this research. One of the major obstacles was the complexity of analysing malware for malware detection. This research used cyber threat intelligence data that contains results on malware analysis such as Cyber Threat Intelligence Program (CTIP). The use of external cyber threat intelligence data helps to avoid wasting resources and redundant work by using the malware data that has been shared by cyber security practitioners, researchers or organisations. The usage of cyber threat intelligence data is practical for a small organisation with limited human and technology capacity. Hence, cyber threat intelligence is included in the improved malware detection framework to implement automated malware analysis. In addition to that, it can be noted that building a malware detection capability and technology is expensive. Therefore the knowledge, as well as the skills to build own tools or integrating established tools are essential. This is also another challenge for the research to use open-source tools and resources in developing the improved malware detection framework. Hence, based on these challenges, the scope of the research are as follows:

i. The research focus was to formulate the malware detection framework that defines the steps to perform malware detection. The machine learning component is only used to automate the malware analysis process of generating information to identify the malware IOCs from the cyber threat intelligence data.

ii. The experiment uses the Cyber Threat Intelligence Program (CTIP) data from Microsoft that contains network data of Botnets incidents in Malaysia.

iii. Data collection is only a method used to obtain the CTIP data and real-time or the near real time data collection is not covered in this research. Solving the real-time or near-real time data collection issue is not part of the research objectives.

iv. This research selects the clustering algorithm based on the algorithm suitability with the cyber threat intelligence data collected. A comparison of K-Means and Gaussian Mixture Model is shown in this research to support the selection of the clustering algorithm.

v. The threat hunting experiment performed on three malware is used to demonstrate the detection of malware. The experiment results are discussed to evaluate if the malware is detected using the framework.

1.7 Research Contribution and Significance

The use of a single technique is unable to handle sophisticated malware (Nguyen et al., 2018) and the traditional signature based approach fails to detect malware (Sibi Chakkaravarthy et al., 2019). This is because sophisticated malware signatures are unknown which make them undetected by existing malware detection solutions e.g. the Intrusion Detection System (IDS), Intrusion Prevention System (IPS) and antivirus (Gandotra et al., 2014). These security systems have limited analysis ability as they only detect or block malware. Sophisticated malware has various ways to hide their malicious intention that aims to evade security systems (Karim et al., 2014). Example of a method to hide malware malicious intention is the metaphoric technique that changes the malware code and creates a new malware signature pattern which makes analysing malware difficult (Alam et al., 2014). This technique implementation complicates malware analysis which further delays malware detection. Malware detection is delayed if malware is analysed manually to discover their signatures (Egele et al., 2012). This creates a predicament as malware cannot be analysed manually one by one to extract the malware indicator of compromise (IOC) e.g. malware signature from large malware samples (Choi et al., 2012).

Therefore, it is evident that as the number of new malware increases, the workload to find new malware signatures increase. This research proposed an improved malware detection framework which defines the guideline in dealing with huge malware sample. The malware detection framework design includes the malware detection solution components as well as the detection approach used. In addition to that, this research provides a better understanding of the underlying malware signature evolution for cyber security practitioners to conduct threat hunting within their IT systems as a proactive mitigation effort. The technical guidelines included in the framework presented is practical and reviewed by professionals as well as academicians in the cyber security field. It is helpful for cyber security practitioners to have a framework for malware detection to provide the best practices for malware detection as the procedures and techniques are crossed-referenced, tested and scientifically proven.

At the present, the number of malware continues to increase as previously shown in Figure 1.1. There are technical frameworks that are available for reference in conducting malware detection such as those proposed in the NIST guideline. NIST provides a general guideline in dealing with malware incidents through the incident response life cycle (NIST, 2013). However, the NIST general guidelines do not go into detail on the implementation process as the life cycle only provides the best practices. This is because a typical Standard Operating Procedures (SOP) with technical guidelines is self-developed by cyber security organisations to conduct malware detection.

Each technique has its advantages and disadvantages. Therefore, a solution such a step by step approach on the tools to use, what to look for and what to do with any suspicious file encountered (Verma et al., 2013) would contribute to the community and is better in helping to establish an understanding of the process to detect malware. Therefore, techniques are combined in a framework to improve the advantages and to decrease the disadvantages (Elhadi et al., 2012). Combining methods and techniques in a framework for malware detection emphasises on making decisions based on data analysis and information evaluation (Foroughi & Luksch, 2018). One of the frameworks proposed in the work of other researchers is B-DAD framework (Elgendy & Elragal, 2016). The framework starts with the intelligence phase that collects data and the second phase called the design phase that analyses the data collected. The third phase is the choice phase that evaluates the analysis result and the framework ends with the implementation phase that operationalises the results. The operationalisation of the result implements the data-driven decision making process to use the data collected and provide information that assists in making decisions. There is an urgent need by cyber security community to advise and provide a technical framework in performing automated malware analysis to detect malware. Therefore, the objectives of this research are to identify the solution for the problem with malware detection, to design a malware detection framework as well as to test the solution for malware detection.

The main contributions of the thesis are the design and implementation of an improved malware detection framework. Additional technical discussions of the malware detection system development and experiments conducted shown are examples for a better understanding of the proposed malware detection framework.

1.8 Thesis Structure

The thesis is divided into six chapters. The following sections provide an overview of each chapter.

1.8.1 Chapter 2 Overview

Chapter 2 provides a detailed review of the literature to obtain relevant information regarding malware detection solutions. The review of malware detection methods used by other researchers provide insights on how to derive an appropriate solution for malware detection. The research works on malware detection are reviewed thoroughly to understand the problems highlighted by cyber security researchers as well as the proposed malware detection components included in the solutions. At the end of this chapter, a summarised review of the malware detection solution and the research gap identified presented.

1.8.2 Chapter 3 Overview

Chapter 3 provides the research design and the plan to achieve the research objectives based on the findings from Chapter 2. The three stages of research design are elaborated in this chapter to show how the research was conducted. At the end of this chapter, the research outcomes are presented along with the research objectives and the research design.

1.8.3 Chapter 4 Overview

Chapter 4 elaborates the Stage 2 of the research work on the framework formulation. The formulation of the improved malware detection framework is based on the second research objective. The improved malware detection framework is formulated based on the identified malware detection solution components is discussed. At the end of this chapter, the formulation of the malware detection framework is presented.

1.8.4 Chapter 5 Overview

Chapter 5 explains the Stage 3 of the research work on implementing and evaluating the formulated framework. An experiment was conducted for the completion of the third research objective, which is to evaluate the improved malware detection framework by performing a threat hunting experiment. The result of implementing the automated malware analysis is used in the formulated framework evaluation. The results include a comparison conducted between UML 1 (K-Means) and UML 2 (GMM) to show the suitability of the machine learning algorithm with the data that affects the output. At the end of the chapter, the detection of three malware is demonstrated in the experiment to explain how the result from the clustering algorithm used assists in detecting malware.

1.8.5 Chapter 6 Overview

Chapter 6 revisits the overall research works conducted in this study that starts from addressing the research problem, research findings obtained and the contributions achieved in this study. Besides stating the research outcomes, this chapter discusses methods employed to achieve the research objectives. In addition to that, the discussion also includes the contributions and significance of the research. At the end of the chapter, the future works of the research are reviewed to understand the possible improvements that can be implemented in future studies.

1.9 Summary

The research background provides an overview of the malware detection current issue. The problem is cyber security practitioners are having difficulty to manually perform signature-based malware detection due to the increasing number of malware. Therefore, the objectives of this research aim to solve this problem by proposing an improved malware detection framework. Review of literature regarding malware detection is thoroughly presented in Chapter 2 to gain more information on the issue and to obtain further knowledge on malware detection solution components.

REFERENCES

Aboughadareh, S., Csallner, C., & Azarmi, M. (2014). Mixed-Mode Malware and Its Analysis. *PPREW '14 (Program Protection and Reverse Engineering Workshop)*, *2*, 1–12. https://doi.org/10.1145/2689702.2689703

Abu, M. S., Selamat, S. R., Ariffin, A., & Yusof, R. (2018). Cyber threat intelligence – Issue and challenges. *Indonesian Journal of Electrical Engineering and Computer Science*, *10*(1), 371–379. https://doi.org/10.11591/ijeecs.v10.i1.pp371-379

Abu, M. S., Selamat, S. R., Yusof, R., & Ariffin, A. (2018). *An Enhancement of Cyber Threat Intelligence Framework Comparative Study of Cyber Threat Intelligence Framework Framework An Enhancement of Cyber Threat Intelligence*. *November*.

Ahmadi, M., Ulyanov, D., Semenov, S., Trofimov, M., & Giacinto, G. (2016). Novel Feature Extraction, Selection and Fusion for Effective Malware Family Classification. *Proceedings of the Sixth ACM on Conference on Data and Application Security and Privacy - CODASPY '16*, 183–194. https://doi.org/10.1145/2857705.2857713

Ahmed, M. E., Nepal, S., & Kim, H. (2018). MEDUSA: Malware detection using statistical analysis of system's behavior. *Proceedings - 4th IEEE International Conference on Collaboration and Internet Computing, CIC 2018*, 272–278. https://doi.org/10.1109/CIC.2018.00044

Ahn, S. H., Kim, N. U., & Chung, T. M. (2014). Big data analysis system concept for detecting unknown attacks. *International Conference on Advanced Communication Technology, ICACT*, 269–272. https://doi.org/10.1109/ICACT.2014.6778962

Ajay, A. K., & C.D., J. (2018). Automated multi-level malware detection system based on reconstructed semantic view of executables using machine learning techniques at VMM. *Future Generation Computer Systems*, *79*, 431–446. https://doi.org/10.1016/j.future.2017.06.002

Ajay Kumara, M. A., & Jaidhar, C. D. (2017). Leveraging virtual machine introspection with memory forensics to detect and characterize unknown malware using machine learning techniques at hypervisor. *Digital Investigation*, *23*, 99–123. https://doi.org/10.1016/j.diin.2017.10.004

Akomolafe, D. T., Ilori, A. O., & Mary, Y. O. (2018). Establishing a process for cyber related crimes investigation through digital forensics. *American Journal of Computer Science and Information Engineering*, *5*(1), 9–14. https://doi.org/https://www.researchgate.net/profile/Abiola_Ilori/publication/327802817_Establishing_a_Process_for_Cyber_Related_Crimes_Investigation_Through_Digital_Forensics_Email_address_Citation/links/5ba55bc045851574f7dd18fb/Establishing-a-Process-for-Cyber-Related-Crimes-Investigation-Through-Digital-Forensics-Email-address-Citation.pdf

Alam, S., Traore, I., & Sogukpinar, I. (2014). Current Trends and the Future of Metamorphic Malware Detection. *Proceedings of the 7th International Conference on Security of Information and Networks - SIN '14*, 411–416. https://doi.org/10.1145/2659651.2659670

Ali Mirza, Q. K., Awan, I., & Younas, M. (2018). CloudIntell: An intelligent malware detection system. *Future Generation Computer Systems*, *86*, 1042–1053. https://doi.org/10.1016/j.future.2017.07.016

Almarri, S., & Sant, P. (2014). Optimised Malware Detection in Digital Forensics. *International Journal of Network Security & Its Applications*, *6*(1), 01–15. https://doi.org/10.5121/ijnsa.2014.6101

Almohannadi, H., Awan, I., Al Hamar, J., Cullen, A., Disso, J. P., & Armitage, L. (2018). Cyber threat intelligence from honeypot data using elasticsearch. *Proceedings - International Conference on Advanced Information Networking and Applications, AINA*, *2018-May*(September), 900–906. https://doi.org/10.1109/AINA.2018.00132

Ariffin, A., D'Orazio, C., Choo, K. K. R., & Slay, J. (2013). IOS forensics: How can we recover deleted image files with timestamp in a forensically sound manner? *Proceedings - 2013 International Conference on Availability, Reliability and Security, ARES 2013*, *2008*, 375–382. https://doi.org/10.1109/ARES.2013.50

Ariyaluran Habeeb, R. A., Nasaruddin, F., Gani, A., Targio Hashem, I. A., Ahmed, E., & Imran, M. (2018). Real-time big data processing for anomaly detection: A Survey. *International Journal of Information Management*, *February*, 1–19. https://doi.org/10.1016/j.ijinfomgt.2018.08.006

Awan, S., & Saqib, N. A. (2016). Detection of malicious executables using static and dynamic features of portable executable (PE) file. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *10067 LNCS*(4), 48–58. https://doi.org/10.1007/978-3-319-49145-5_6

Bat-Erdene, M., Park, H., Li, H., Lee, H., & Choi, M. (2017). Entropy analysis to classify unknown packing algorithms for malware detection. *International Journal of Information Security*, *16*(3), 227–248. https://doi.org/10.1007/s10207-016-0330-4

Belaoued, M., & Mazouzi, S. (2015). Towards an Automatic Method for API Association Extraction for PE-Malware Categorization. *Proceedings of the International Conference on Intelligent Information Processing, Security and Advanced Communication - IPAC '15*, 1–4. https://doi.org/10.1145/2816839.2816921

Bhatt, P., Yano, E. T., & Gustavsson, P. (2014). Towards a framework to detect multi-stage advanced persistent threats attacks. *Proceedings - IEEE 8th International Symposium on Service Oriented System Engineering, SOSE 2014*, *June 2015*, 390–395. https://doi.org/10.1109/SOSE.2014.53

Biggio, B., Rieck, K., Ariu, D., Wressnegger, C., Corona, I., Giacinto, G., & Roli, F. (2014). Poisoning behavioral malware clustering. *Proceedings of the ACM Conference on Computer and Communications Security*, *2014-Novem*(November), 27–36. https://doi.org/10.1145/2666652.2666666

Boom, C., Continue, T., & Motivated, T. A. (2019). McAfee Labs Threats Report: December 2018. *Computer Fraud & Security*, *2019*(1), 4. https://doi.org/10.1016/s1361-3723(19)30004-1

Branco, R. R., & Shamir, U. (2010). Architecture for automation of malware analysis. *Proceedings of the 5th IEEE International Conference on Malicious and Unwanted Software, Malware 2010*, 106–112. https://doi.org/10.1109/MALWARE.2010.5665786

Brynjolfsson, E., & McElheran, K. (2016). The rapid adoption of data-driven decision-making. *American Economic Review*, *106*(5), 133–139. https://doi.org/10.1257/aer.p20161016

Bulazel, A., & Yener, B. (2017). A Survey On Automated Dynamic Malware Analysis Evasion and Counter-Evasion. *Proceedings of the 1st Reversing and Offensive-Oriented Trends Symposium on   - ROOTS*, 1–21.

https://doi.org/10.1145/3150376.3150378

Burnap, P., French, R., Turner, F., & Jones, K. (2018). Malware classification using self organising feature maps and machine activity data. *Computers and Security*, *73*, 399–410. https://doi.org/10.1016/j.cose.2017.11.016

Cabau, G., Buhu, M., & Oprisa, C. P. (2017). Malware classification based on dynamic behavior. *Proceedings - 18th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, SYNASC 2016*, 315–318. https://doi.org/10.1109/SYNASC.2016.057

Cakir, B., & Dogdu, E. (2018). Malware classification using deep learning methods. *Proceedings of the ACMSE 2018 Conference on - ACMSE '18*, 1–5. https://doi.org/10.1145/3190645.3190692

Chantzios, T., Koloveas, P., Skiadopoulos, S., Kolokotronis, N., Tryfonopoulos, C., Bilali, V. G., & Kavallieros, D. (2019). The quest for the appropriate cyber-threat intelligence sharing platform. *DATA 2019 - Proceedings of the 8th International Conference on Data Science, Technology and Applications*, 369–376. https://doi.org/10.5220/0007978103690376

Choi, Y. H., Han, B. J., Bae, B. C., Oh, H. G., & Sohn, K. W. (2012). Toward Extracting Malware Features for Classification using Static and Dynamic Analysis. *International Conference on Computing and Networking Technology*, 126–129.

Cohen, A., & Nissim, N. (2018). Trusted detection of ransomware in a private cloud using machine learning methods leveraging meta-features from volatile memory. *Expert Systems with Applications*, *102*, 158–178. https://doi.org/10.1016/j.eswa.2018.02.039

Crandall, J. R., Ensafi, R., Forrest, S., Ladau, J., & Shebaro, B. (2008). The ecology of Malware. *Proc. New Security Paradigms Workshop 2008*, 99–106. https://doi.org/10.1145/1595676.1595692

Dan Lo, C. T., Ordóñez, P., & Cepeda, C. (2016). Feature selection and improving classification performance for malware detection. *Proceedings - 2016 IEEE International Conferences on Big Data and Cloud Computing, BDCloud 2016, Social Computing and Networking, SocialCom 2016 and Sustainable Computing and Communications, SustainCom 2016*, 560–566. https://doi.org/10.1109/BDCloud-SocialCom-SustainCom.2016.87

David, O. E., & Netanyahu, N. S. (2015). DeepSign: Deep learning for automatic malware signature generation and classification. *2015 International Joint Conference on Neural Networks (IJCNN)*, 1–8. https://doi.org/10.1109/IJCNN.2015.7280815

Deckert, A., & Sarre, R. (2017). CYBERCRIME IN AUSTRALIA. *The Palgrave Handbook of Australian and New Zealand Criminology, Crime and Justice*, *April*, 1–916. https://doi.org/10.1007/978-3-319-55747-2

Deli, M. S. M., Ismail, S. A., Kama, N., Yusop, O. M., Azmi, A., & Yahya, Y. (2018). Malware log files for Internet investigation using hadoop: A review. *2017 IEEE Conference on Big Data and Analytics, ICBDA 2017*, *2018-Janua*, 87–92. https://doi.org/10.1109/ICBDAA.2017.8284112

Divita, J., & Hallman, R. A. (2017). An Approach to Botnet Malware Detection Using Nonparametric Bayesian Methods. *Proceedings of the 12th International Conference on Availability, Reliability and Security - ARES '17*, 1–9. https://doi.org/10.1145/3098954.3107010

Egele, M., Scholte, T., Kirda, E., & Kruegel, C. (2012). A survey on automated dynamic malware-analysis techniques and tools. *ACM Computing Surveys*, *44*(2), 1–42. https://doi.org/10.1145/2089125.2089126

Elgendy, N., & Elragal, A. (2016). Big Data Analytics in Support of the Decision Making Process. *Procedia Computer Science*, *100*, 1071–1084. https://doi.org/10.1016/j.procs.2016.09.251

Elhadi, A. A. E., Maarof, M. A., & Osman, A. H. (2012). Malware detection based on hybrid signature behavior application programming interface call graph. *American Journal of Applied Sciences*, *9*(3), 283–288. https://doi.org/10.3844/ajassp.2012.283.288

Elyas, M., Maynard, S. B., Ahmad, A., & Lonie, A. (2014). Towards a systemic framework for digital forensic readiness. *Journal of Computer Information Systems*, *54*(3), 97–105. https://doi.org/10.1080/08874417.2014.11645708

Fan, Y., Ye, Y., & Chen, L. (2016). Malicious sequential pattern mining for automatic malware detection. *Expert Systems with Applications*, *52*, 16–25. https://doi.org/10.1016/j.eswa.2016.01.002

Feng, Z., Xiong, S., Cao, D., Deng, X., Wang, X., Yang, Y., Zhou, X., Huang, Y., & Wu, G. (2015). HRS A Hybrid Framework for malware detection. *Proceedings of the 2015 ACM International Workshop on International Workshop on Security and Privacy Analytics - IWSPA '15*, *10*, 19–26.

https://doi.org/10.1145/2713579.2713585

Firdausi, I., Lim, C., Erwin, A., & Nugroho, A. S. (2010). Analysis of Machine learning Techniques Used in Behavior-Based Malware Detection. *2010 Second International Conference on Advances in Computing, Control, and Telecommunication Technologies*, 201–203. https://doi.org/10.1109/ACT.2010.33

Foroughi, F., & Luksch, P. (2018). Data Science Methodology for Cybersecurity Projects. *Computer Science & Information Technology*, *3*(1), 01–14. https://doi.org/10.5121/csit.2018.80401

Galal, H. S., Mahdy, Y. B., & Atiea, M. A. (2016). Behavior-based features model for malware detection. *Journal of Computer Virology and Hacking Techniques*, *12*(2), 59–67. https://doi.org/10.1007/s11416-015-0244-0

Gandotra, E., Bansal, D., & Sofat, S. (2014). Integrated Framework for Classification of Malwares. *Proceedings of the 7th International Conference on Security of Information and Networks - SIN '14*, 417–422. https://doi.org/10.1145/2659651.2659738

Garcia-Serrano, A. (2015). *Anomaly Detection for malware identification using Hardware Performance Counters*. 1–12. http://arxiv.org/abs/1508.07482

Gengenbach, M., Chassanoff, A., & Olsen, P. (2012). Integrating digital forensics into born-digital workflows: The BitCurator project. *Proceedings of the ASIST Annual Meeting*, *49*(1). https://doi.org/10.1002/meet.14504901343

Ghalaty, N. F., & Ben Salem, M. (2019). A Hierarchical Framework to Detect Targeted Attacks using Deep Neural Network. *Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018*, 5021–5026. https://doi.org/10.1109/BigData.2018.8622131

Ghate, A. M., & Malik, L. G. (2015). Survey on designing framework for analyzing twitter spammers using forensic method. *2015 International Conference on Pervasive Computing: Advance Communication Technology and Application for Society, ICPC 2015*, *00*(c), 1–4. https://doi.org/10.1109/PERVASIVE.2015.7087158

Ghiasi, M., Sami, A., & Salehi, Z. (2015). Dynamic VSA: a framework for malware detection based on register contents. *Engineering Applications of Artificial Intelligence*, *44*, 111–122. https://doi.org/10.1016/j.engappai.2015.05.008

Gill, B., Coffee-Borden, B., & Hallgren, K. (2014). A Conceptual Framework for

Data-Driven Decision Making. *Mathematica Policy Research Reports*, *609*. https://doi.org/10.1080/19397030802591196

Gunjan, V. K., Kumar, A., & Avdhanam, S. (2013). A survey of cyber crime in India. *2013 15th International Conference on Advanced Computing Technologies, ICACT 2013*, *May 2019*. https://doi.org/10.1109/ICACT.2013.6710503

Han, K. S., Kang, B., & Im, E. G. (2011). Malware classification using instruction frequencies. *Proceedings of the 2011 ACM Symposium on Research in Applied Computation - RACS '11*, 298. https://doi.org/10.1145/2103380.2103441

Hassen, M., & Chan, P. K. (2017). Scalable Function Call Graph-based Malware Classification. *Proceedings of the Seventh ACM on Conference on Data and Application Security and Privacy - CODASPY '17*, 239–248. https://doi.org/10.1145/3029806.3029824

Hopkins, M., & Dehghantanha, A. (2016). Exploit Kits: The production line of the Cybercrime economy? *2015 2nd International Conference on Information Security and Cyber Forensics, InfoSec 2015*, *2015*(January 2015), 23–27. https://doi.org/10.1109/InfoSec.2015.7435501

Huang, H. De, Lee, C. S., Wang, M. H., & Kao, H. Y. (2014). IT2FS-based ontology with soft-computing mechanism for malware behavior analysis. *Soft Computing*, *18*(2), 267–284. https://doi.org/10.1007/s00500-013-1056-0

Huda, S., Abawajy, J., Alazab, M., Abdollalihian, M., Islam, R., & Yearwood, J. (2016). Hybrids of support vector machine wrapper and filter based framework for malware detection. *Future Generation Computer Systems*, *55*, 376–390. https://doi.org/10.1016/j.future.2014.06.001

Huda, S., Islam, R., Abawajy, J., Yearwood, J., Hassan, M. M., & Fortino, G. (2018). A hybrid-multi filter-wrapper framework to identify run-time behaviour for fast malware detection. *Future Generation Computer Systems*, *83*, 193–207. https://doi.org/10.1016/j.future.2017.12.037

Huda, S., Miah, S., Mehedi Hassan, M., Islam, R., Yearwood, J., Alrubaian, M., & Almogren, A. (2017). Defending unknown attacks on cyber-physical systems by semi-supervised approach and available unlabeled data. *Information Sciences*, *379*, 211–228. https://doi.org/10.1016/j.ins.2016.09.041

Incer, I., Theodorides, M., Afroz, S., & Wagner, D. (2018). Adversarially Robust Malware Detection Using Monotonic Classification. *Proceedings of the Fourth*

ACM International Workshop on Security and Privacy Analytics, 54–63. https://doi.org/10.1145/3180445.3180449

Intelligence, K. T. (n.d.). *Kaspersky Threat Intelligence*.

Kang, B., Yerima, S., McLaughlin, K., & Sezer, S. (2015). PageRank in malware categorization. *Proceedings of the 2015 Conference on Research in Adaptive and Convergent Systems - RACS*, 291–295. https://doi.org/10.1145/2811411.2811514

Karim, A., Salleh, R. Bin, Shiraz, M., Shah, S. A. A., Awan, I., & Anuar, N. B. (2014). Botnet detection techniques: review, future trends, and issues. *Journal of Zhejiang University SCIENCE C*, *15*(11), 943–983. https://doi.org/10.1631/jzus.C1300242

Kaspersky. (2018). *Incident Response Analytics Report 2018*.

Katzir, Z., & Elovici, Y. (2018). Quantifying the resilience of machine learning classifiers used for cyber security. *Expert Systems with Applications*, *92*, 419–429. https://doi.org/10.1016/j.eswa.2017.09.053

Kesavan, K., Vimukthi Bannakkotuwa, S., Wickramanayake, V. V. Y., De Silva, M. P. D. H., Fernando, J. M. D., Sampath, K. K. K. K., & Rupasinghe, L. (2016). *ClusterMal: Automated Malware Analysis with clustering, anomaly detection and classification of existing and new behavioral analysis*. *August*.

Khitan, S. J., Hadi, A., & Atoum, J. (2017). PDF Forensic Analysis System using YARA. *International Journal of Computer Science and Network Security*, *17*(5), 77–85.

Kim, S., Kim, T., & Im, E. G. (2013). Real-time malware detection framework in intrusion detection systems. *Proceedings of the 2013 Research in Adaptive and Convergent Systems on - RACS '13*, 351–352. https://doi.org/10.1145/2513228.2513297

Kim, T., Park, J. Bin, Cho, I. G., Kang, B., Im, E. G., & Kang, S. (2014). Similarity calculation method for user-define functions to detect malware variants. *Proceedings of the 2014 Conference on Research in Adaptive and Convergent Systems - RACS '14*, 236–241. https://doi.org/10.1145/2663761.2664222

Kiwia, D., Dehghantanha, A., Choo, K. K. R., & Slaughter, J. (2018). A cyber kill chain based taxonomy of banking Trojans for evolutionary computational intelligence. *Journal of Computational Science*, *27*, 394–409. https://doi.org/10.1016/j.jocs.2017.10.020

Kozik, R. (2018). Distributing extreme learning machines with Apache Spark for NetFlow-based malware activity detection. *Pattern Recognition Letters*, *101*, 14–20. https://doi.org/10.1016/j.patrec.2017.11.004

Kumar, A., Kuppusamy, K. S., & Aghila, G. (2017). A learning model to detect maliciousness of portable executable using integrated feature set. *Journal of King Saud University - Computer and Information Sciences*. https://doi.org/10.1016/j.jksuci.2017.01.003

Kumar, R., & Vaishakh, A. R. E. (2016). Detection of Obfuscation in Java Malware. *Physics Procedia*, *78*(December 2015), 521–529. https://doi.org/10.1016/j.procs.2016.02.097

Kumar, V., Kumar, S., & Gupta, A. K. (2016). Real-time Detection of Botnet Behavior in Cloud Using Domain Generation Algorithm. *Proceedings of the International Conference on Advances in Information Communication Technology & Computing - AICTC '16*, 1–3. https://doi.org/10.1145/2979779.2979848

Kwon, I., & Im, E. G. (2017). Extracting the Representative API Call Patterns of Malware Families Using Recurrent Neural Network. *Proceedings of the International Conference on Research in Adaptive and Convergent Systems - RACS '17*, 202–207. https://doi.org/10.1145/3129676.3129712

Lai, C., Lu, C., & Lee, H. (2018). Implementation of Adversarial Scenario to Malware Analytic. *Proceedings of the 2nd International Conference on Machine Learning and Soft Computing - ICMLSC '18*, 127–132. https://doi.org/10.1145/3184066.3184078

Lee, Y., Lee, J., & Soh, W. (2018). Trend of Malware Detection Using Deep Learning. *Proceedings of the 2nd International Conference on Education and Multimedia Technology - ICEMT 2018*, 102–106. https://doi.org/10.1145/3206129.3239430

Lim, C., & Ramli, K. (2015). Mal-ONE: A unified framework for fast and efficient malware detection. *Proceedings of 2014 2nd International Conference on Technology, Informatics, Management, Engineering and Environment, TIME-E 2014*, 1–6. https://doi.org/10.1109/TIME-E.2014.7011581

Liu, L., Wang, B., Yu, B., & Zhong, Q. (2017). Automatic malware classification and new malware detection using machine learning. *Frontiers of Information Technology & Electronic Engineering*, *18*(9), 1336–1347.

https://doi.org/10.1631/FITEE.1601325

Liu, W., Ren, P., Liu, K., & Duan, H. X. (2011). Behavior-based malware analysis and detection. *Proceedings - 2011 1st International Workshop on Complexity and Data Mining, IWCDM 2011*, *60203044*, 39–42. https://doi.org/10.1109/IWCDM.2011.17

Louk, M., Lim, H., Lee, H., & Atiquzzaman, M. (2015). An effective framework of behavior detection-advanced static analysis for malware detection. *14th International Symposium on Communications and Information Technologies, ISCIT 2014*, 361–365. https://doi.org/10.1109/ISCIT.2014.7011932

Luthfi, A., & Prayudi, Y. (2016). Process Model of Digital Forensics Readiness Scheme (DFRS) as a Recommendation of Digital Evidence Preservation. *Proceedings - 4th International Conference on Cyber Security, Cyber Warfare, and Digital Forensics, CyberSec 2015*, *June 2016*, 117–122. https://doi.org/10.1109/CyberSec.2015.31

Mao, W., Cai, Z., Yang, Y., Shi, X., & Guan, X. (2018). From big data to knowledge: A spatio-temporal approach to malware detection. *Computers and Security*, *74*, 167–183. https://doi.org/10.1016/j.cose.2017.12.005

Martinelli, F., Mercaldo, F., Nardone, V., & Santone, A. (2017). Malware and Formal Methods. *Proceedings of the 12th International Conference on Availability, Reliability and Security - ARES '17*, 1–6. https://doi.org/10.1145/3098954.3107012

Martinelli, F., Mercaldo, F., Nardone, V., Santone, A., Sangaiah, A. K., & Cimitile, A. (2018). Evaluating model checking for cyber threats code obfuscation identification. *Journal of Parallel and Distributed Computing*, *119*, 203–218. https://doi.org/10.1016/j.jpdc.2018.04.008

Masabo, E., Kaawaase, K. S., & Sansa-Otim, J. (2018). Big Data: Deep Learning for Detecting Malware. *Proceedings of the 2018 International Conference on Software Engineering in Africa*, *i*, 20–26. https://doi.org/10.1145/3195528.3195533

Mayhew, M., Atighetchi, M., Adler, A., & Greenstadt, R. (2015). Use of machine learning in big data analytics for insider threat detection. *Proceedings - IEEE Military Communications Conference MILCOM*, *2015-Decem*(September), 915–922. https://doi.org/10.1109/MILCOM.2015.7357562

Mckemmish, R. (1999). What is Forensic Computing ? *Change*, *118*(118), 1–6.

http://www.mendeley.com/catalog/forensic-computing-2/

Moftah, R. A., Maatuk, A. M., Plasmann, P., & Aljawarneh, S. (2015). An overview about the polymorphic worms signatures. *ACM International Conference Proceeding Series*, *24-26-Sept*. https://doi.org/10.1145/2832987.2833031

Mohaisen, A., Alrawi, O., & Mohaisen, M. (2015). AMAL: High-fidelity, behavior-based automated malware analysis and classification. *Computers and Security*, *52*, 251–266. https://doi.org/10.1016/j.cose.2015.04.001

Moustafa, N., Creech, G., & Slay, J. (2017). *Big Data Analytics for Intrusion Detection System: Statistical Decision-Making Using Finite Dirichlet Mixture Models* (Issue May, pp. 127–156). https://doi.org/10.1007/978-3-319-59439-2_5

Muller, S., Lancrenon, J., Harpes, C., Le Traon, Y., Gombault, S., & Bonnin, J. M. (2018). A training-resistant anomaly detection system. *Computers and Security*, *76*, 1–11. https://doi.org/10.1016/j.cose.2018.02.015

Naval, S., Laxmi, V., Gaur, M. S., Raja, S., Rajarajan, M., & Conti, M. (2015). Environment–Reactive Malware Behavior: Detection and Categorization. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 8872, pp. 167–182). https://doi.org/10.1007/978-3-319-17016-9_11

Naval, S., Laxmi, V., Rajarajan, M., Gaur, M. S., & Conti, M. (2015). Employing Program Semantics for Malware Detection. *IEEE Transactions on Information Forensics and Security*, *10*(12), 2591–2604. https://doi.org/10.1109/TIFS.2015.2469253

Nguyen, M. H., Nguyen, D. Le, Nguyen, X. M., & Quan, T. T. (2018). Auto-detection of sophisticated malware using lazy-binding control flow graph and deep learning. *Computers and Security*, *76*, 128–155. https://doi.org/10.1016/j.cose.2018.02.006

Nissim, N., Cohen, A., Glezer, C., & Elovici, Y. (2015). Detection of malicious PDF files and directions for enhancements: A state-of-the art survey. *Computers and Security*, *48*, 246–266. https://doi.org/10.1016/j.cose.2014.10.014

Nissim, N., Moskovitch, R., Rokach, L., & Elovici, Y. (2014). Novel active learning methods for enhanced PC malware detection in windows OS. *Expert Systems with Applications*, *41*(13), 5843–5857. https://doi.org/10.1016/j.eswa.2014.02.053

NIST. (2013). *NIST Special Publication 800-83 Guide to Malware Incident Prevention and Handling for Desktops and Laptops NIST Special Publication 800-83 Guide to Malware Incident Prevention and Handling for Desktops and Laptops*. 7. https://doi.org/10.6028/NIST.SP.800-83r1

Nunes, E., Diab, A., Gunn, A., Marin, E., Mishra, V., Paliath, V., Robertson, J., Shakarian, J., Thart, A., & Shakarian, P. (2016). Darknet and deepnet mining for proactive cybersecurity threat intelligence. *IEEE International Conference on Intelligence and Security Informatics: Cybersecurity and Big Data, ISI 2016*, 7–12. https://doi.org/10.1109/ISI.2016.7745435

Okereke, A. O., & Chukwunonso, C. E. (2018). Malware Analysis and Mitigation in Information Preservation. *Journal of Computer Engineering*, *20*(4), 53–62. https://doi.org/10.9790/0661-2004015362

Pai, S., Troia, F. Di, Visaggio, C. A., Austin, T. H., & Stamp, M. (2017). Clustering for malware classification. *Journal of Computer Virology and Hacking Techniques*, *13*(2), 95–107. https://doi.org/10.1007/s11416-016-0265-3

Pandey, A. K., Tripathi, A. K., Kapil, G., Singh, V., Khan, M. W., Agrawal, A., Kumar, R., & Khan, R. A. (2020). *Trends in Malware Attacks*. *January*, 47–60. https://doi.org/10.4018/978-1-7998-1558-7.ch004

Payne, B. D., Carbone, M., Sharif, M., & Lee, W. (2008). Lares: An architecture for secure active monitoring using virtualization. *Proceedings - IEEE Symposium on Security and Privacy*, *May*, 233–247. https://doi.org/10.1109/SP.2008.24

Pektaş, A., & Acarman, T. (2017). Classification of malware families based on runtime behaviors. *Journal of Information Security and Applications*, *37*, 91–100. https://doi.org/10.1016/j.jisa.2017.10.005

Pirscoveanu, R. S., Hansen, S. S., Larsen, T. M. T., Stevanovic, M., Pedersen, J. M., & Czech, A. (2015). Analysis of malware behavior: Type classification using machine learning. *2015 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA)*, 1–7. https://doi.org/10.1109/CyberSA.2015.7166128

Pluskal, O. (2015). Behavioural malware detection using efficient SVM implementation. *Proceedings of the 2015 Conference on Research in Adaptive and Convergent Systems - RACS*, 296–301. https://doi.org/10.1145/2811411.2811516

Provost, F., & Fawcett, T. (2013). Data Science and its Relationship to Big Data and

Data-Driven Decision Making. *Big Data*, *1*(1), 51–59. https://doi.org/10.1089/big.2013.1508

Raff, E., Zak, R., Cox, R., Sylvester, J., Yacci, P., Ward, R., Tracy, A., McLean, M., & Nicholas, C. (2018). An investigation of byte n-gram features for malware classification. *Journal of Computer Virology and Hacking Techniques*, *14*(1), 1–20. https://doi.org/10.1007/s11416-016-0283-1

Raphel, J., & Vinod, P. (2017). Heterogeneous Opcode Space for Metamorphic Malware Detection. *Arabian Journal for Science and Engineering*, *42*(2), 537–558. https://doi.org/10.1007/s13369-016-2264-6

Raphel, J., & Vinod, P. (2015). Information theoretic method for classification of packed and encoded files. *Proceedings of the 8th International Conference on Security of Information and Networks - SIN '15*, *52*(11), 296–303. https://doi.org/10.1145/2799979.2800015

Rasmi, M., & Al Qerem, A. (2015). PNFEA: A Proposal Approach for Proactive Network Forensics Evidence Analysis to Resolve Cyber Crimes. *International Journal of Computer Network and Information Security*, *7*(2), 25–32. https://doi.org/10.5815/ijcnis.2015.02.03

Rathore, H., Agarwal, S., Sahay, S. K., & Sewak, M. (2018). Malware detection using machine learning and deep learning. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11297 LNCS*, 402–411. https://doi.org/10.1007/978-3-030-04780-1_28

Raval, U. R., & Jani, C. (2015). Implementing and Improvisation of K-means Clustering. *International Journal of Computer Science and Mobile Computing*, *4*(11), 72–76.

Razak, M. F. A., Anuar, N. B., Salleh, R., & Firdaus, A. (2016). The rise of "malware": Bibliometric analysis of malware study. In *Journal of Network and Computer Applications* (Vol. 75, pp. 58–76). Elsevier. https://doi.org/10.1016/j.jnca.2016.08.022

Rieck, K., Holz, T., Willems, C., Düssel, P., & Laskov, P. (2008). Learning and Classification of Malware Behavior. In U. Flegel, E. Markatos, & W. Robertson (Eds.), *Detection of Intrusions and Malware, and Vulnerability Assessment* (Vol. 7591, Issue July, pp. 108–125). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-70542-0_6

Sahay, S. K., & Sharma, A. (2016). Grouping the Executables to Detect Malwares with High Accuracy. *Physics Procedia*, *78*(December 2015), 667–674. https://doi.org/10.1016/j.procs.2016.02.115

Saini, A., Gandotra, E., Bansal, D., & Sofat, S. (2014). Classification of PE Files using Static Analysis. *Proceedings of the 7th International Conference on Security of Information and Networks - SIN '14*, 429–433. https://doi.org/10.1145/2659651.2659679

Saleem, S., Popov, O., & Bagilli, I. (2014). Extended abstract digital forensics model with preservation and protection as umbrella principles. *Procedia Computer Science*, *35*(C), 812–821. https://doi.org/10.1016/j.procs.2014.08.246

Samtani, S., Abate, M., Benjamin, V., & Li, W. (2019). The Palgrave Handbook of International Cybercrime and Cyberdeviance. *The Palgrave Handbook of International Cybercrime and Cyberdeviance*, *September*. https://doi.org/10.1007/978-3-319-90307-1

Satrya, G. B., Cahyani, N. D. W., & Andreta, R. F. (2015). The Detection of 8 Type Malware botnet using Hybrid Malware Analysis in Executable File Windows Operating Systems. *Proceedings of the 17th International Conference on Electronic Commerce 2015 - ICEC '15*, 1–4. https://doi.org/10.1145/2781562.2781567

Scofield, D., Miles, C., & Kuhn, S. (2017). Fast Model Learning for the Detection of Malicious Digital Documents. *Proceedings of the 7th Software Security, Protection, and Reverse Engineering / Software Security and Protection Workshop*, 3:1--3:8. https://doi.org/10.1145/3151137.3151142

Sharma, S., Rama Krishna, C., & Sahay, S. K. (2019). Detection of advanced malware by machine learning techniques. *Advances in Intelligent Systems and Computing*, *742*, 333–342. https://doi.org/10.1007/978-981-13-0589-4_31

Shijo, P. V., & Salim, A. (2015). Integrated static and dynamic analysis for malware detection. *Procedia Computer Science*, *46*(Icict 2014), 804–811. https://doi.org/10.1016/j.procs.2015.02.149

Shrestha, P., Maharjan, S., de la Rosa, G. R., Sprague, A., Solorio, T., & Warner, G. (2014). Using string information for malware family identification. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8864*, 686–697. https://doi.org/10.1007/978-3-319-12027-0 55

Sibi Chakkaravarthy, S., Sangeetha, D., Vaidehi, V., Chakkaravarthy, S. S., Sangeetha, D., & Vaidehi, V. (2019). A Survey on malware analysis and mitigation techniques. *Computer Science Review*, *32*, 1–23. https://doi.org/S1574013718301114

Simou, S., Kalloniatis, C., Kavakli, E., & Gritzalis, S. (2014). Cloud Forensics: Identifying the Major Issues and Challenges. In M. Jarke, J. Mylopoulos, C. Quix, C. Rolland, Y. Manolopoulos, H. Mouratidis, & J. Horkoff (Eds.), *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 8484, Issue June, pp. 271–284). Springer International Publishing. https://doi.org/10.1007/978-3-319-07881-6_19

Singh, K., Guntuku, S. C., Thakur, A., & Hota, C. (2014). Big Data Analytics framework for Peer-to-Peer Botnet detection using Random Forests. *Information Sciences*, *278*, 488–497. https://doi.org/10.1016/j.ins.2014.03.066

Singh, V., & Verma, N. K. (2019). *An Entropy-based Variable Feature Weighted Fuzzy k-Means Algorithm for High Dimensional Data*. 1–7. http://arxiv.org/abs/1912.11209

Song, L., Huang, H., Zhou, W., Wu, W., & Zhang, Y. (2016). Learning from Big Malwares. *Proceedings of the 7th ACM SIGOPS Asia-Pacific Workshop on Systems - APSys '16*, 1–8. https://doi.org/10.1145/2967360.2967367

Staheli, D., Mancuso, V., Harnasch, R., Fulcher, C., Chmielinski, M., Kearns, A., Kelly, S., & Vuksani, E. (2016). Collaborative Data Analysis And Discovery For Cyber Security. *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*, *August*, 1–8. https://www.usenix.org/conference/soups2016/workshop-program/wsiw16/presentation/staheli%5Cnhttps://www.researchgate.net/publication/306264470_Collaborative_Data_Analysis_and_Discovery_for_Cyber_Security%5Cnhttps://www.researchgate.net/profile/Vincent_Man

Tedre, M., & Moisseinen, N. (2014). *Experiments in Computing : A Survey*. *2014*. https://doi.org/10.1155/2014/549398

Tounsi, W., & Rais, H. (2018). A survey on technical threat intelligence in the age of sophisticated cyber attacks. *Computers and Security*, *72*(November), 212–233. https://doi.org/10.1016/j.cose.2017.09.001

Ucci, D., Aniello, L., & Baldoni, R. (2017). *Survey on the Usage of Machine*

Learning Techniques for Malware Analysis. http://arxiv.org/abs/1710.08189

Ullah, F., & Ali, M. (2018). Architectural Tactics for Big Data Cybersecurity Analytic Systems : A Review. *Arxiv.Org*, 1–48. http://arxiv.org/abs/1802.03178%0Ahttps://arxiv.org/abs/1802.03178

Venkatraman, S., & Alazab, M. (2018). Classification of Malware Using Visualisation of Similarity Matrices. *Proceedings - 2017 Cybersecurity and Cyberforensics Conference, CCC 2017*, *2018-Septe*, 3–8. https://doi.org/10.1109/CCC.2017.11

Verma, A., M.S., R., A.K., G., Jeberson, W., & Singh, V. (2013). A Literature Review on Malware and its Analysis. *International Journal of Current Research and Review*.

VirusTotal. (n.d.). *Searching for file scan reports*. Retrieved August 1, 2019, from https://support.virustotal.com/hc/en-us/articles/115002739245-Searching

Wang, J., Xue, Y., Liu, Y., & Tan, T. H. (2015). JSDC: A hybrid approach for JavaScript malware detection and classification. *ASIACCS 2015 - Proceedings of the 10th ACM Symposium on Information, Computer and Communications Security*, *April*, 109–120. https://doi.org/10.1145/2714576.2714620

Wang, L., Wang, G., & Alexander, C. A. (2015). Big Data and Visualization: Methods, Challenges and Technology Progress. *Digital Technologies*, *1*(1), 33–38. https://doi.org/10.12691/dt-1-1-7

Wheelus, C., Bou-Harb, E., & Zhu, X. (2016). Towards a big data architecture for facilitating cyber threat intelligence. *2016 8th IFIP International Conference on New Technologies, Mobility and Security, NTMS 2016*, *October*. https://doi.org/10.1109/NTMS.2016.7792484

Xue, Y., Wang, J., Liu, Y., Xiao, H., Sun, J., & Chandramohan, M. (2015). Detection and classification of malicious JavaScript via attack behavior modelling. *Proceedings of the 2015 International Symposium on Software Testing and Analysis - ISSTA 2015*, *1*, 48–59. https://doi.org/10.1145/2771783.2771814

Yan, W., & Wu, E. (2009). Toward automatic discovery of malware signature for anti-virus cloud computing. *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering*, *4 LNICST*(PART 1), 724–728. https://doi.org/10.1007/978-3-642-02466-5_70

Ye, Y., Li, T., Adjeroh, D., & Iyengar, S. S. (2017). A Survey on Malware Detection Using Data Mining Techniques. *ACM Computing Surveys*, *50*(3), 1–40.

https://doi.org/10.1145/3073559

Yeo, M., Koo, Y., Yoon, Y., Hwang, T., Ryu, J., Song, J., & Park, C. (2018). Flow-based malware detection using convolutional neural network. *International Conference on Information Networking*, *2018-Janua*, 910–913. https://doi.org/10.1109/ICOIN.2018.8343255

Yewale, A., & Singh, M. (2017). Malware detection based on opcode frequency. *Proceedings of 2016 International Conference on Advanced Communication Control and Computing Technologies, ICACCCT 2016*, *978*, 646–649. https://doi.org/10.1109/ICACCCT.2016.7831719

You, I., & Yim, K. (2010). Malware obfuscation techniques: A brief survey. *Proceedings - 2010 International Conference on Broadband, Wireless Computing Communication and Applications, BWCCA 2010*, 297–300. https://doi.org/10.1109/BWCCA.2010.85

Zhao, H., Li, M., Wu, T., & Yang, F. (2018). Evaluation of supervised machine learning techniques for dynamic malware detection. *International Journal of Computational Intelligence Systems*, *11*(1), 1153–1169. https://doi.org/10.2991/ijcis.11.1.87

Zhou, D., Yan, Z., Fu, Y., & Yao, Z. (2018). A survey on network data collection. In *Journal of Network and Computer Applications* (Vol. 116, Issue April, pp. 9–23). Elsevier Ltd. https://doi.org/10.1016/j.jnca.2018.05.004

Zhou, Y., & Jiang, X. (2012). Dissecting Android malware: Characterization and evolution. *Proceedings - IEEE Symposium on Security and Privacy*, *4*, 95–109. https://doi.org/10.1109/SP.2012.16

Zolkipli, M. F., & Jantan, A. (2011). An approach for malware behavior identification and classification. *ICCRD2011 - 2011 3rd International Conference on Computer Research and Development*, *1*, 191–194. https://doi.org/10.1109/ICCRD.2011.5764001

Appendix A Create Index Pseudo Code

```
<?php
require 'vendor/autoload.php';
use Elasticsearch\ClientBuilder;
$client = ClientBuilder::create()->build();
$params = [
'index' => 'ctip_index',
'body' => [
'settings' => [
'number_of_shards' => 3,
'number_of_replicas' => 2
],
'mappings' => [
//mapping configuration here
]
]
]
]
];
$response = $client->indices()->create($params);
?>
```

Appendix B Indexing Pseudo Code

```
<?php
require 'vendor/autoload.php';
use Elasticsearch\ClientBuilder;
$client = ClientBuilder::create()->build();
//Read the data from directory
//Data parsing
$params = [
'index' => 'ctip_index',
'type' => 'threat_log',
'body' => [
//Match data attributes to index attributes
]
];
$response = $client->index($params);
?>
```

Appendix C Malware Heatmap

Appendix D First Conference 2019 Presentation

Appendix E Secure Conference 2019 Paper

AHMAD NAIM IRFAN <ahmad.naim.irfan@graduate.utm.my>

SC 2019 notification for paper 10

3 messages

SC 2019 <sc2019@easychair.org> Wed, May 15, 2019 at 4:36 PM

To: Ahmad Naim Irfan Aswami Fadillah <ahmad.naim.irfan@graduate.utm.my>

Dear Author(s)

Congratulations, the review process for the International Industry-Researchers Conference on Cyber Security 2019 (SecureConf2019) has been completed. The conference received numerous papers from different countries, which were reviewed by international experts, and a number of papers have been selected for the presentation and publication.

Based on the recommendations of the reviewers and the Program Committee, we are pleased to inform you that your paper identified above has been accepted for oral presentation and publication. You are cordially invited to deliver an oral presentation at Secure Conf 2019 that will be held by July 3 - 4, 2019 at School of Computing and IT, Lakeside Campus, Taylor's University, Subang Jaya Malaysia. Please visit the conference website (https://sc2019.wsconferences.com/) for the registration process and further details.

Important: you must complete the following steps (within two weeks of receiving this acceptance email) while preparing the camera-ready version of your paper.

1. Carefully revise your paper based on the reviewers' comments.
2. Please note selected papers will be published with PERTANIKA Journal of Science Technology (ISSN 0128-7680), Scopus Index after suggested editing/extending the papers as per the journal standards and policies.
3. Format your camera-ready paper according to the template carefully, using the PERTANIKA Journal of Science Technology (ISSN 0128-7680) (http://www.pertanika.upm.edu.my/JST.php), please follow the instructions of the journal provided at (http://www.pertanika.upm.edu.my/instructions_to_authors.php).
4. Complete the list of authors along with their affiliation.
5. Provide the copyright form (Form will be forwarded to you, after camera-ready version submission).
6. Send your Camera-Ready papers through Easy Chair by re-uploading it in (.doc) and in (.pdf) formats.
7. Send the Correction Form which includes the corrections made as per the reviewer's comments to noorzaman.jhanjhi@taylors.edu.my with title a Correction Form of (paper ID).
8. Thank you for submitting a paper to SecureConf 2019. We look forward to meeting you at the School of Computing and IT SoCIT, Lakeside Campus, Taylor's University, Subang Jaya, Malaysia.

Best Regards

Noor Zaman Jhanjhi Ph.D.

Conference Co-Chair, SecureConf 2019

https://sc2019.wsconferences.com/

SUBMISSION: 10

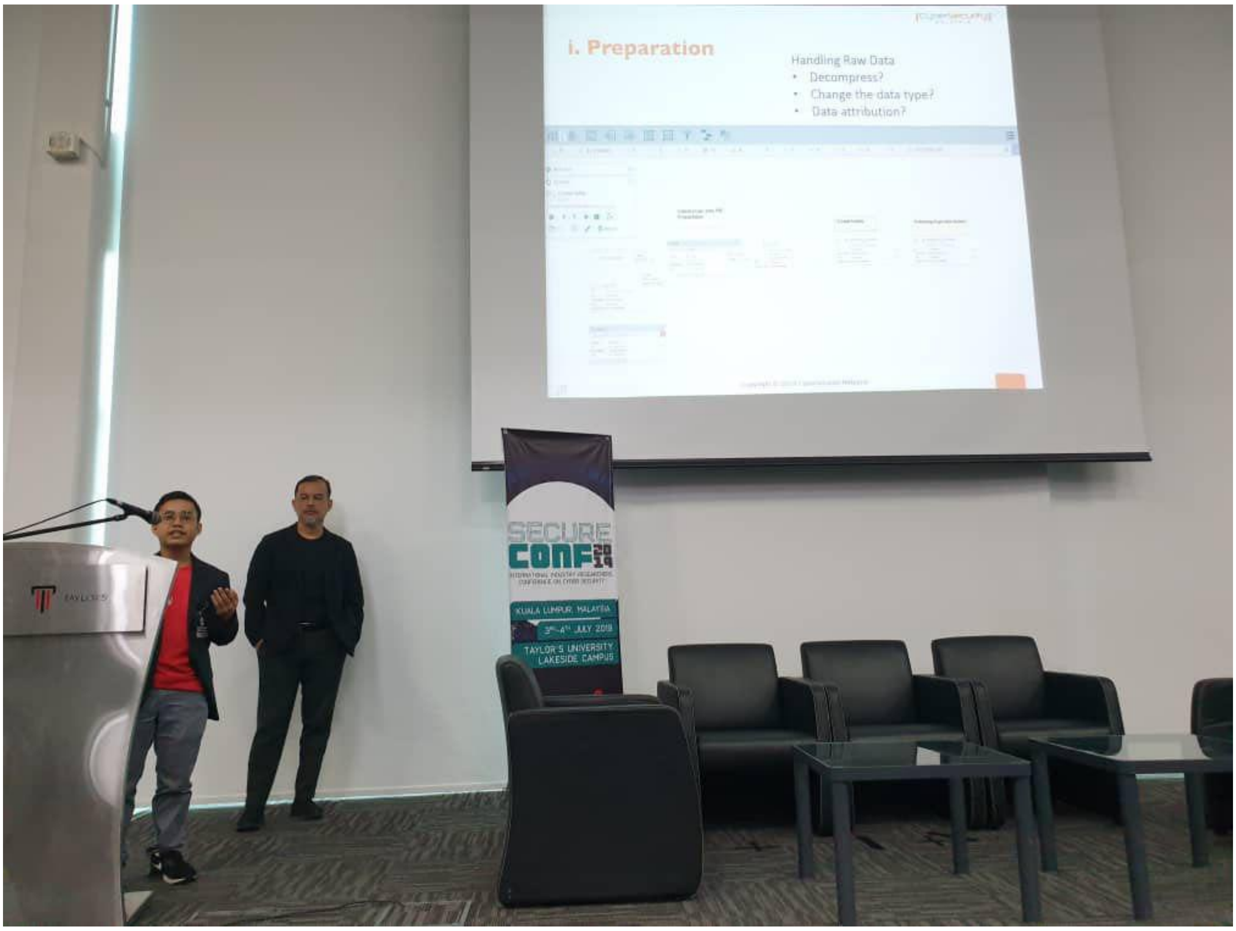
TITLE: MALWARE FORENSIC DETECTION FRAMEWORK USING BIG DATA SYSTEM

Appendix F Secure Conference 2019 Schedule

Conference Day 2, 4th July 2019, Thursday

Time	Agenda
09:00 - 09:30	Keynote Address 2 by Prof. Emeritus Sureswaran Ramadass Chairman, IPv6 & IOT Centre of Expertise, International Telecommunication Union (ITU) "Botnets: Rise of the Dark Dragon"
09;30-10;00	Taylor's University Industry Advisory Panel - MACROKIOSK Bhd Mr Patrick Hew, Chief Technology Officer "Gaining consumers trust online "
10:00-10:30	Mr. Philip Victor, Director, Cyber8Lab, Australia "Preparing and Defending Against Cyber Attacks: A Strategic Approach"
10:30-11:00	Tea Break
11:00-11:30	"Ts Dr Aswami Fadillah Bin Mohd Ariffin, Senior VP & Digital Forensics Scientist, CyberSecurity Malaysia (CSM) Ahmad Naim Irfan, Postgraduate Student, University Teknologi Malaysia (UTM) "Malware Forensic Detection Framework Using Big Data System'
11:30-12:00	Major Dr.Pramod Gurubasappa Bagali, CEO, Witty Charman CoTS Sdn Bhd " Review of Fast Healthcare Interoperability Resource (FHIR) standards for Information Security of Healthcare Data"

Appendix G Secure Conference 2019 Presentation



Appendix H Secure Conference 2019 Certification


SECURE CONF 2019

WIRAHSWASTA

CERTIFICATION OF RECOGNITION

Presented to

Ahmad Naim Irfan

For being Paper Presenter

International Industry-Researchers Conference on Cybersecurity 2019

3rd-4th July 2019

Shaan Bin Sivagurunathan
CEO
Wirahswasta Sdn Bhd

Appendix I ICSCA 2020 Paper

ICSCA 2020

Notification of Acceptance

February 18-21, 2020 | Langkawi, Malaysia

www.icsca.org

Dear Ahmad Naim Irfan

Congratulations! The review process for 2020 9th International Conference on Software and Computer Applications (ICSCA 2020) has been completed. The conference received 58 submissions from nearly 15 different countries and regions, which were reviewed by international experts, and about 37 papers have been selected for presentation and publication. Based on the recommendations of the reviewers and the Technical Program Committee, we are pleased to inform you that your paper has been accepted for publication and presentation. You are cordially invited to present the paper on ICSCA 2020 to be held during **February 18-21, 2020** in **Langkawi, Malaysia** (**Click**).

Paper ID :	A26
Paper Title :	A Malware Detection Framework Based on Forensic and Unsupervised Machine Learning Methodologies

will be published in International Conference Proceedings Series by ACM, which will be archived in the ACM Digital Library, and indexed by Ei Compendex and Scopus and submitted to be reviewed by Thomson Reuters Conference Proceedings Citation Index (ISI Web of Science). ISBN of ICSCA 2019 Conference Proceedings: 978-1-4503-5414-1

Good News: ICSCA2017, 2018, 2019 conference proceedings have been archieved into ACM digital library and indexed by Scopus and Ei Compendex 5 months after conference.

Finally, we would like to further extend our congratulations to you and we are looking forward to meeting you in Langkawi, Malaysia!

Yours sincerely,

ICSCA International Conference on Software and Computer Applications

ICSCA 2020 Organizing Committee

icsca_general@163.com
Langkawi, Malaysia

LIST OF PUBLICATIONS & PRESENTATIONS

No	Publication and Presentation
1	The framework is part of the research presented at First Conference 2019 Appendix D
2	Paper entitled "Malware forensic detection framework using big data system" is accepted at Secure Conference 2019 Appendix E (Paper Withdrawn)
3	Presented the paper entitled "Malware forensic detection framework using big data system" at Secure Conference 2019 , Taylors College , 4th July 2019 Appendix F, Appendix G and Appendix H
4	Paper entitled "A Malware detection framework based on forensic and unsupervised machine learning methodologies" is accepted at ICSCA 2020 Appendix I
5	To present the paper entitled "A Malware detection framework based on forensic and unsupervised machine learning methodologies" at ICSCA 2020, Bella Vista Waterfront Langkawi, February 2020 Appendix I