UNIVERSITI PUTRA MALAYSIA

A DYNAMIC MALWARE DETECTION IN CLOUD PLATFORM

NANI LEE YER FUI

FSKTM 2019 41

A DYNAMIC MALWARE DETECTION IN CLOUD PLATFORM

By

NANI LEE YER FUI

Thesis submitted to the School of Graduate Studies,

Universiti Putra Malaysia,

in Fulfilment of the Requirements for the Master of Information Security

JUN 2019

© COPYRIGHT UPM

COPYRIGHT PAGE

All material contained within the thesis, including without limitation text, logos, icons, photographs and all other artworks, is copyright material of Universiti Putra Malaysia unless otherwise stated. Use may be made of any material contained within the thesis for non-commercial purposes from the copyright holder. Commercial use of material may only be made with the express, prior, written permission of Universiti Putra Malaysia.

Copyright © Universiti Putra Malaysia

© COPYRIGHT UPM

ABSTRACT

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Master of Information Security

A DYNAMIC MALWARE DETECTION IN CLOUD PLATFORM

By

NANI LEE YER FUI

JUNE 2019

Supervisor: Dr Aziah Bte Asmawi

Faculty: Faculty of Computer Science and Information Technology

Abstract:

Cloud Computing platform is the practices of remote manage network resources such as storage, application hosted on the internet rather than physical server or personal computer. Hence cloud computing not only provides high availability on elastic resources, scalable and cost efficient. This is why this this platform is widely used in information technology (IT) to support technology infrastructure and services.

However, due to the complexity environment and scalability of services, one of a highest security issue is malware attacks; where some of the antivirus scanner unable to detect metamorphic malware or encrypted malware where these kind of malware

© COPYRIGHT UPM

able to bypass some traditional protection solution. This is why a high recognition rate and a good precision detection are important to eliminate high false positive rate.

Machine learning (ML) classifiers are critical role in the artificial intelligent-system such as medical assistance detect whether the cell is cancerous or benign or to convert the spoken audio file into a text file. However machine learning will require learn from high amplitude of input data; classify then only able to generate a reliable model with high detection rate.

The objective in this work is to study and performs detection based on dynamic malware analysis and classification is through WEKA classifier and Random Forest Jupyter Notebook. In this work we assess five classifiers, for instance the Random Forest in WEKA, Decision Tree (J48) in WEKA and Bayes Network (BN) in WEKA tool, and Random Forest in Jupyter Notebook comprised 9600 malware dataset obtained from Kaggle to exhibit the model's effectiveness, out of which additional 600 are new malware dataset, whereby previous solution consist 9000 malware dataset.

© COPYRIGHT UPM

ABSTRAK

Abstrak tesis yang dikemukan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Sarjana Keselamatan Maklumat

DETEKSI PERISIAN HASAD DI PENGKOMPUTRAAN AWAN

Oleh

NANI LEE YER FUI

JUN 2019

Penyelia: Dr Aziah Bte Asmawi

Fakulti: Fakulti Sains Komputer dan Teknologi Maklumat

Abstrak:

Pengkomputeran Awan menguruskan sumber rangkaian jauh seperti storan, aplikasi tersedia yang dihoskan di internet dan bukan di pelayan fizikal atau komputer peribadi. Oleh itu, pengkomputeran awan bukan sahaja menyediakan ketersediaan tinggi pada sumber elastik, berskala dan kos yang cekap. Inilah sebabnya mengapa platform ini digunakan secara meluas dalam teknologi maklumat (IT) untuk menyokong infrastruktur tknologi dan perkhidmatan.

Walau bagaimanapun, disebabkan oleh persekitaran yang rumit dan kebolehskalan perkhidmatan, salah satu isu keselamatan tertinggi adalah serangan malware; di mana beberapa pengimbas antivirus tidak dapat mengesan malware metamorfik atau malware yang disulitkan (encrypted) di mana jenis malware ini dapat memintas

© COPYRIGHT UPM

beberapa penyelesaian perlindungan tradisional. Inilah sebabnya mengapa kadar pengiktirafan yang tinggi dan pengesanan ketepatan yang baik adalah penting untuk menghapuskan kadar positif palsu yang tinggi.

Pengklasifikasian pembelajaran mesin memainkan peranan penting dalam sistem pintar(artificial intelligent-system)seperti bantuan perubatan mengesan sama ada sel kanser atau bukan sel kanser, serta untuk menukar fail audio yang dituturkan ke dalam fail teks. Walau bagaimanapun, pembelajaran mesin memerlukan pembelajaran dari data input yang banyak; klasifikasi barulah dapat menghasilkan model yang boleh dipercayai dengan kadar pengesanan yang tepat.

Objektif utama dalam kerja ini adalah mengkaji dan melakukan pengesanan berdasarkan analisis malware dinamik dan klasifikasi adalah melalui pengelas perisian *WEKA* dan Random Fores dari Notebook Jupiter. Dalam kerja ini kami menilai lima pengelas, contohnya *Random Forest* dari *WEKA*, *Decision Tree* (J48) dari *WEKA* and *Bayes Network* (BN) dari WEKA tool, beserta dengan *Random Forest* dari *Jupyter Notebook* dimana ia mengandungi 9600(sembilan ribu enam ratus) dataset malware yang diperoleh dari *Kaggle* untuk mempamerkan model keberkesanan, di mana tambahan 600(enam ratus) adalah dataset malware baru, di mana penyelesaian sebelumnya terdiri daripada 9000(Sembilan ribu) dataset malware.

© COPYRIGHT UPM

ACKNOWLEDGEMENT

First of all, I would like to acknowledge everyone during the course of this Master Of Information Security program, to my supervisor Dr. Aziah Asmawi, who helped me endlessly since the proposal phase until the end of completing this thesis, thank you for sacrifice your family time help me to get through.

And to the entire lecturers in Master of Information Security program, the valuable knowledge and patience imparted to me has been a great asset to apply in our daily work.

To my family, friends and course mates' thank you for making this journey wonderful, unforgettable and memorable.

© COPYRIGHT UPM

APPROVAL

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Master of Information Security. The members of the Supervisory Committee were as follows:

Signature:_____

DR AZIAH BINTI ASMAWI

Faculty of Computer Science and Information Technology

Universiti Putra Malaysia

(Supervisor)

Date:

© COPYRIGHT UPM

DECLARATION

Declaration by graduate student

I hereby confirm that:

- this thesis is my original work;
- quotations, illustration and citation have been duly referenced;
- this thesis has not been submitted previously or concurrently for any other degree at any other institutions;
- intellectual property from the thesis and copyright of thesis are fully-owned by Universiti Putra Malaysia, as according to the Universiti Putra Malaysia (Research) Rules 2012;
- written permission must be obtained from supervisor and the office of Deputy Vice-Chancellor (Research and Innovation) before thesis is published (in the form of written, printed or in electronic form) including books, journals, report, lecturer notes, learning modules or any other materials as stated in the Universiti Putra Malaysia (Research) Rules 2012;
- there is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to the Universiti Putra Malaysia(Graduate Studies) Rules 2003 (Revision 2012-2013) and the Universiti Putra Malaysia (Research) Rules 2012.

Signature: _____Date: _____

Name and Matric No.: **NANI LEE YER FUI**

GS49541

© COPYRIGHT UPM

TABLE OF CONTENTS

		Page
ABSTRACT		ii
ABSTRAK		iv
ACKNOWLEDGEMENT		vi
APPROVAL		vii
DECLARATION		viii
TABLE OF CONTENTS		ix
LIST OF TABLES		xii
LIST OF FIGURES		xiii
CHAPTER		
1	**INTRODUCTION**	1
	1.1 BACKGROUND	1
	1.2 PROBLEM STATEMENT	2
	1.3 RESEARCH OBJECTIVE	2
	1.4 RESEARCH SCOPE	2
	1.5 THESIS STRUCTURE	3
2	**BACKGROUND AND RELATED WORK**	4
	2.1 BACKGROUND	4
	2.2 CLOUD COMPUTING: AN OVERVIEW	4

© COPYRIGHT UPM

	2.3	DEPLOYMENT MODEL FOR CLOUD COMPUTING	5
		2.2.1 Service Model For Cloud Computing	6
		2.2.2 The Treacherous 12 Cloud Computing	8
		2.2.3 Cloud Security Breaches by Malware	10
	2.4	COMMON MALWARE TYPES	10
		2.4.1 Cyber Attacks Statistics	13
		2.4.2 Differences Between Static Malware Analysis and dynamic malware analysis	14
		2.4.3 Static Malware Analysis	14
		2.4.4 Dynamic Malware Analysis	14
	2.5	RELATED WORK	15
	2.6	CONCLUSION	22
3		**RESEARCH METHODLOGY**	23
	3.1	OVERVIEW	23
	3.2	RESEARCH FRAMEWORK	23
		3.2.1 Detection and Classification	23
		3.2.2 Malware and Clean Ware Dataset	24
		3.2.3 Cuckoo Automated Malware Analysis	24
		3.2.4 Random Forest	25
		3.2.5 WEKA Software	25
		3.2.6 Machine Learning Classifier	25
		3.2.7 Naming And Calculation	25
	3.3	SYSTEM SETUP	26
	3.4	SOFTWARE AND HARDWARE REQUIREMENTS	27
	3.5	CONCLUSION	27
4		**RESULT AND DISCUSSION**	28
	4.1	RESULT AND ANALYSIS	28
		4.1.1 Result From Random Forest WEKA	28
	4.2	COMPARISON OF RESULTS	30
		4.2.1 Results and Dataset by Previous Work	31
	4.3	CONCLUSION	35

© COPYRIGHT UPM

5	**CONCLUSION AND RECOMMENDATION**	36
	5.1 CONCLUSION	36
	5.2 LIMITATIONS	36
	5.3 RECOMMENDATIONS FOR FUTURE RESEARCH	37
REFERENCES		38
APPENDIX A		40
APPENDIX B		43

© COPYRIGHT UPM

LIST OF TABLES

Table No.

2.0	Summary of Deployment Model For Cloud Computing.	6
2.1	Service Model Characteristics.	7
3.0	Naming and Calculation.	25
3.1	Software and Hardware requirements.	26
4.0	Results in WEKA Algorithm.	28
4.1	Random Forest Jupyter Notebook.	29
4.2	Comparison Of Processing Time (In Seconds) In WEKA Algorithm.	30
4.3	Performance Comparison Of Our Proposed With Previous Solutions.	30
4.4	Algorithm WEKA Detection Of Metamorphic Malware.	32
4.5	Results are describe using TPR, FPR PPV, F-M and AUC.	33
4.6	Performances of malware classification	33

© COPYRIGHT UPM

LIST OF FIGURES

Figure No.

2.1	Deployment Model For Cloud Computing.	6
2.2	Service Model For Cloud Computing.	7
2.3	Motivation Behind Attacks January 2018.	13
2.4	2018-Cyber-Attacks-Statistics.	13
2.5	Comparison Static Analysis And Dynamic Analysis.	15
2.6	System Methodology Using Deep Belief Network (DBN).	16
2.7	An Overview of Detection Classification Metamorphic malware.	17
2.8	Overview Of Detection scheme in cloud architecture.	19
2.9	Overall System Flow For Both Malware Detection And Classification.	20
2.9.1	Mechanism Module.	21
2.9.2	System Flow For Detection Technique.	22
3.1	Overview Of System Flow.	23
3.2	Illustration of The Framework.	24
3.3	Illustration Of The System Setup.	26

© COPYRIGHT UPM

LIST OF FIGURES

Figure No.

4.1	Comparison Performance Of Our Proposed Method Against anchor paper.	29
4.2	Comparison Performance Of Our Proposed With Previous Solutions	30
4.3	Analysis results.	32
4.4	Analysis results.	33
4.5	Accuracy of malware detection.	34
4.6	Training Time.	34
4.7	Time taken to output a class.	35

© COPYRIGHT UPM

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND

The objective of cloud computing is to provide shared services such as servers, applications by multiple users pay per use to minimize the cost of ownership. This attack can be ranging from internal to external threat such phishing attack, malware injection, abuse of cloud services, insider threats.

In addition, other attacks such as shell shock, heart bleed which used by attackers to disable service by perform denial of service or to distribute malware. Hence in order to protect such attacks, effective and efficient malware detection scheme is necessary.

There are two main approaches for malware detection static detection and dynamic detection, in static detection where analysis is analysing the execution code without really execute the code or program and because just collect the file information e.g.: hashing file. Wwhereas dynamic analysis analyse the behaviour of malicious code or program in a contained and isolated virtual environment or sandbox to differentiate whether the code is malicious or benign and not only that also able to identify malicious application which use obfuscation techniques such as packing or code encryption or polymorphism.

© COPYRIGHT UPM

1.2 PROBLEM STATEMENT

Some malware scanners are unable to detect malware that specifically uses metamorphism and polymorphism. Mainly because signature files do not match the profile signature in the database or malware signature enable malware easily slip though using code obfuscation.

In this work, we re-implement algorithm dynamic malware analysis to resolve the metamorphism and polymorphism malware, with the high detection rate which involves capture the feature vectors from each process such as API, URL for analysis to decide whether the executable file is a malicious or benign file (clean ware).

1.3 RESEARCH OBJECTIVE

The objective in this work is to re-implement a Dynamic Malware Detection Mechanism scheme in Cloud Platform to resolve the problem of metamorphic malware or encrypted malware by enhancing the detection accuracy. There are four machine learning classifiers were evaluated namely Random Forest in WEKA, J-48 in WEKA, naïve Bayes in WEKA and also Random Forest Jupyter Notebook.

1.4 RESEARCH SCOPE

The scope of in this work is to improve detection accuracy of the malware detection algorithm in cloud platform. To perform the security analysis of the proposed solution to test the strength of the algorithm to ensure the accuracy of classification.

© COPYRIGHT UPM

1.5 THESIS STRUCTURE

This thesis is structured as per below:

Chapter 1: Introduction

This chapter presents the introduction to the Cloud Computing, security issues in cloud computing, objective, study, research problem of this study, proposed method, research questions, objectives followed by research scope.

Chapter 2: Background and related work

This chapter presents the background of cloud computing, security issues, types of malware, difference between static and dynamic malware and related work to this research.

Chapter 3: Methodology

This chapter presents a system of method and activity used in this work, which includes data collection method, type of dataset, classification and result analysis will discuss in this chapter.

Chapter 4: Result and discussion

This chapter presents the result of the analysis from this work and to result comparison between this work and previous work.

Chapter 5: Conclusion and Recommendation

This chapter presents the conclusion from this work and recommendation for future work.

© COPYRIGHT UPM

REFERENCES

[1] Title: What is cloud computing, Url: https://azure.microsoft.com/en-us/overview/what-is-cloud-computing/.Accessed: 19-10-2018.

[2] Title: The dirty dozen: 12 top cloud security threats for 2018, Url: https://www.csoonline.com/article/3043030/security/12-top-cloud-security-threatsfor2018.html.Accessed: 19-10-2018.

[3] Title: Biggest hacks so far, Url: https://www.wired.com/story/2017-biggest Hacks-so-far/.Accessed: 20-10-2018.

[4] Title: Malware types and classifications. Url: https://www.lastline.com/blog/malware-types-and-classifications.Accessed: 20-10-2018.

[5] Title: January 2018 cyber-attacks statistics.Url:https://www.hackmageddon.com/2018/02/22/january-2018 cyber-attacks statistics/.Accessed: 19-10-2018.

[6] Title: Difference Between Static malware analysis and dynamic malware analysis.Url:http://www.differencebetween.net/technology/differencebetween-static-malware-analysis-and-dynamic-malware analysis/Accessed: 20-10-2018.

[7] Eli (Omid) David and Nathan S. Netanyahu. (2015). *DeepSign: Deep Learning for Automatic Malware Signature Generation and Classification*. International Joint Conference on Neural Networks (IJCNN), pages 1–8, Killarney, Ireland, July 2015.

[8] Vishakha Mehra, Vinesh Jain, Dolly Uppal. (2015). D*aCoMM: Detection and Classification of Metamorphic Malware*. 2015 Fifth International Conference on Communication Systems and Network Technologies.

[9] Shiva Darshan S.L.1, Ajay Kumara M.A.2, and Jaidhar C.D. (2016).*Windows Malware Detection Based on Cuckoo Sandbox Generated Report Using Machine LearningAlgorithm.* 2016 11th International Conference On Industrial and Information Systems (ICIIS).

[10] Michael R. Watson, Noor-ul-hassan Shirazi, Angelos K. Marnerides, Andreas Mauthe, and David Hutchison. (2016). *Malware Detection in Cloud Computing Infrastructures*. IEEE Transactions On Dependable And Secure Computing, Vol. 13, No. 2, March/April 2016.

[11] Steven Strandlund Hansen, Thor Mark Tampus Larsen, Matija Stevanovic and Jens Myrup Pedersen. (2016). *An Approach for Detection and Family Classification of Malware Based on Behavioral Analysis. 2016* International Conference on Computing, Networking and Communications (ICNC), Workshop on Computing, Networking and Communications (CNC).

© COPYRIGHT UPM

[12] Wei YIN, Hongjian ZHOU, Mingyang WANG, Zhiwen JIN, Jun XU. (2018). *A Dynamic Malware Detection Mechanism Based on Deep Learning*. IJCSNS International Journal of Computer Science and Network Security, VOL.18 No.7, July 2018.

[13] Saket Kumar,Chandra Bhim Bhan Singh. (2018). A *zero-day resistant malware detection method for securing Cloud using SVM and SandboxingTechniques*. Proceedings of the 2nd International Conference on Inventive Communication and Computational Technologies (ICICCT 2018). IEEE Xplore Compliant - Part Number: CFP18BAC-ART; ISBN:978-1-5386-1974-2.

[14] Title: Weka 3: Machine Learning Software in Java. Url:https://www.cs.waikato.ac.nz/ml/weka/. Accessed: 20-10-2018.

[15] Title: The Jupyter Notebook. Url:https://jupyter.org/.Accessed: 20-10-2018.

[16] Title: Malware dataset. Url https://www.kaggle.com/c/malware-classification. Accessed: 20-10-2018.

[17] Yunan Zhang, Qingjia Huang, Xinjian Ma, Zeming Yang and Jianguo Jiang (2016). Using Multi-features and Ensemble Learning Method for Imbalanced Malware Classification. 2016 IEEE TrustCom/BigDataSE/ISPA

[18] Title: Cuckoo Automated Malware Analysis. Url: https://cuckoosandbox.org/. Accessed: 20-10-2018.

[19] Title: Understanding Random Forests Classifiers in Python. Url: https://www.datacamp.com/community/tutorials/random-forestsclassifier-python. Accessed: 20-10-2018.

[20] Title: Weka3: Machine Learning Software in Java. Url: https://www.cs.waikato.ac.nz/ml/weka/. Accessed: 20-10-2018.

[21] Title: Data Mining with Weka. Url: https://www.futurelearn.com/courses/data-mininwithweka/0/steps/25384. Accessed: 20-10-2018.

[22] Title: Microsoft Malware Classification Challenge, Url: http://arxiv.org/abs/1802.10135 Accessed : 11-03-2018

© COPYRIGHT UPM