CLOUD WORM DETECTION AND RESPONSE TECHNIQUE BY INTEGRATING THE ENHANCED GENETIC ALGORITHM AND THREAT LEVEL

Hasan Mahmoud Sha'ban Kanaker

Thesis submitted in fulfillment for the degree of
DOCTOR OF PHILOSOPHY
SCIENCE AND TECHNOLOGY

UNIVERSITI SAINS ISLAM MALAYSIA
Nilai

March 2018

BIODATA OF AUTHOR

Hasan Mahmoud Sha'ban Kanaker (Matric N.4120178) from **Jordan** holds **Masterdegree**in*Computer Science* from Al **B**alqa' **A**pplied **U**niversity (**BAU**) and a **Bachelor** degree in ***Computer Information System*** from Al-**Z**aytoonah **U**niversity of**J**ordan (**ZUJ**).

He is currently a Ph.D. Candidate in Science and Technology at **Universiti Sains Islam Malaysia (USIM)**. **Kanaker** has almost **seven years' experience** in Network and System Administrator. Previously he worked as Administrator in the Department of Information Technology at Middle East University, Amman - Jordan. **Kanaker** also worked as Network and System Administrator at Istishari Hospital, Amman - Jordan.

ACKNOWLEDGEMENTS

ALLAH (S.W.T) says in the Quran: "***And whatever of blessings and good things you have, it is from Allah***" [al-Nahl 16:53]. I am deeply grateful to **ALLAH (S.W.T)**, for giving me good health to successfully complete my PhD course, including this dissertation **The Prophet Muhammad** (peace be upon him) said as narrated by Abu Hurairah: "***He who does not thank people, does not thank Allah***" (Ahmad, Tirmidhi).

This dissertation would not have been possible if it were not for the invaluable guidance, support, and encouragement I received.

I would like to express my deep gratitude to **Malaysia**, which has not only given me an environment conducive to pursue my degree but also the kindness, hospitality, and assistance I have received from the **Malaysian people**. Forever, I will pray for **Malaysia** for everlasting prosperity.

I would also like to express my deep appreciation to **UNIVERSITI SAINS ISLAM MALAYSIA (USIM)** and especially to the **Fakulti Sains dan Teknologi (FST)** for not only giving me the opportunity to pursue my PhD degree here but also providing me with the facilities and resources I needed to successfully complete my PhD degree.

My very special appreciation and thanks should go to my principal supervisor **Prof.Madya Dr.Madihah Mohd Saudi**, for her invaluable support, guidance, patience and care. **Prof. MadyaDr. Madihah Mohd Saudi has** always been a source of hope, encouragement, and willpower for me to do my very best. She tremendously assisted me throughout the dissertation journey. **Prof. Madya Dr.Madihah Mohd Saudi**, please receive my heartfelt gratitude; and I will always be deeply grateful to you.

I wish to heartily thank my co-supervisor **Dr. Mohd Fadzli Marhusin**, who has been highly resourceful to the work on this dissertation to the extent that if it were not for his contribution, this dissertation would not be a success.

I am deeply grateful to **Prof.Madya Dr. Zaiton Hassan,** who gave me a Research Methodology course from which I obtained useful skills that made me choose the most appropriate research design for my research. I would like to deeply thank **Prof. Dr. Bhasah abu Bakar,** who gave me data analysis course, from where I gained important skills that were useful in analysing data for this dissertation.

I am extremely grateful to my **Parents**, who have given my unending support throughout my education journey and in my entire life. My **Parents'** care, support, encouragement, and best Dua'a have made my study, including this dissertation, a success. My **Parents** have always been there for me; and have always seized every opportunity to cheer me up all the time. May **ALLAH (S.W.T)** grant my parents health and blessings always.

I would like to thank my **Brothers** and **Sisters** for their love and prayers. My deep thanks go to my **loving wife (Fatimah Hayan Al khani)**, who has travelled with our lovely children (**Mahmoud and Banah)** to **Malaysia** and providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this dissertation.

I would like to thank all the **research participants** for their time and help they gave throughout my research. Their contribution to this dissertation has been central, and I will forever be sincerely grateful to them.

ABSTRACT

A worm is known as a malicious code that can replicate, infect and propagate itself without attaching itself to any host. Unlike other malicious codes, such as a Trojan horse and a virus, a worm can cause serious damage due to its payload and interrupt or stop cloud resources and services which then lead to loss of money, confidential information and productivity for organisations or users that rely on data storage, services and applications that run on the cloud. Detecting and stopping cloud worm attacks have become a hard and challenging endeavour. Therefore, this research proposes a cloud worm classification based on its features. The research also develops a cloud worm detection technique by integrating the enhanced genetic algorithm and proposes a cloud response technique based on threat levels. The proposed cloud worm detection and response technique is evaluated based on accuracy rates. The novelty and strengths of the proposed technique for cloud worm detection and response lies in the worm cloud classification, the integration of the enhancement genetic algorithm and the threat level measurement that impact confidentiality, integrity and availability. For the enhanced genetic algorithm (EGA), new parameters have been constructed for the existing genetic algorithm based on the seleection proportional of fitness, tree mutation ,tree crossover and evolution controller to increase the 's accuracy rate of the cloud worm detection technique. The genetic algorithm is based on the idea of the natural selection process and genetics which imitates the biological process, thereby helping solve unconstrained and constrained optimisation problems. The method of new offspring creation from the fittest parents has been mapped and used to detect unknown future cloud worm attacks. Threat level is measured by security metrics on the basis of confidentiality, integrity and availability impact. An experiment has been conducted in a controlled lab environment based on knowledge discovery techniques using open source tools. This research used 1195 datasets in the experiment wherein dynamic analysis and security metrics have been applied. From the experimental result, the EGA technique has produced an overall detection accuracy rate of 99.749 % with 0.014 % false positive rate, thus outperforming the existing work of the OlexGA technique with 41.05% improvement. This research has produced a technique that can detect and respond to cloud worm attacks.

ABSTRAK

Cecacing dikenal sebagai kod berbahaya yang boleh mereplikasi, menjangkiti dan menyebarkan dirinya tanpa kemunculan dirinya kepada mana-mana hos. Tidak seperti kod berbahaya yang lain, seperti Trojan horse dan virus, cecacing boleh menyebabkan kerosakan yang serius akibat muatan dan mengganggu atau menghentikan sumber dan perkhidmatan yang kemudiannya menyebabkan kehilangan wang, maklumat sulit dan produktiviti untuk organisasi atau pengguna yang bergantung pada penyimpanan data, perkhidmatan dan aplikasi yang dilaksanakan di awan. Mengesan dan menghentikan serangan cecacing awan telah menjadi usaha yang sukar dan mencabar. Oleh itu, kajian ini mencadangkan pengelasan cecacing awan berdasarkan ciri-cirinya. Penyelidikan ini juga membangunkan teknik pengesanan cecacing awan dengan mengintegrasikan algoritma genetik yang dipertingkatkan dan mencadangkan teknik tindak balas awan berdasarkan tahap ancaman. Cadangan pengesanan cecacing awan dan teknik tindak balas dinilai berdasarkan kadar ketepatan. Keberkesanan dan kekuatan teknik yang dicadangkan untuk pengesanan dan tindak balas cecacing awan terletak pada klasifikasi cecacing awan, integrasi algoritma genetik peningkatan dan pengukuran tahap ancaman yang memberi kesan kerahsiaan, integriti dan ketersediaan. Untuk algoritma genetik yang dipertingkatkan (EGA), parameter baru telah dibina untuk algoritma genetik yang sedia ada berdasarkan pemilihan berkadar kecergasan, mutasi pokok dan crossover pokok dan pengawal evolusi untuk meningkatkan kadar ketepatan pengesanan cecacing awan. Algoritma genetik adalah berdasarkan kepada proses pemilihan semulajadi dan genetik yang meniru proses biologi, dengan itu membantu menyelesaikan masalah pengoptimuman yang tidak terkawal dan terkawal. Kaedah penciptaan anak baru dari ibu bapa yang paling tepat telah dipetakan dan digunakan untuk mengesan serangan cecacing awan masa depan yang tidak diketahui. Tahap ancaman diukur oleh metrik keselamatan berdasarkan kerahsiaan, integriti dan kesan ketersediaan. Eksperimen telah dijalankan di persekitaran makmal terkawal berdasarkan teknik penemuan pengetahuan menggunakan alat sumber terbuka. Kajian ini menggunakan 1195 dataset dalam eksperimen di mana analisis dinamik dan metrik keselamatan telah digunakan. Daripada hasil eksperimen, teknik EGA telah menghasilkan kadar ketepatan pengesanan keseluruhan sebanyak 99.749% dengan kadar positif palsu 0.014%, dengan itu mengatasi kerja sedia ada teknik OlexGA dengan peningkatan 41.05%. Penyelidikan ini telah menghasilkan satu teknik yang dapat mengesan dan bertindak balas terhadap serangan cecacing awan.

ملخص

تعرف الدّودة في الحوسبة على أنها نوع من أنواع الرموز الخبيثة، والتي لديها القدرة على تكرار وإصابة ونشر نفسها من دون إضافتها إلى أي مضيف. وبالمقارنة مع الرموز الخبيثة الأخرى، مثل حصان طروادة والفيروس، يمكن للدّودة أن تسبب أضراراً جسيمةً؛ بسبب حمولتها ومقاطعتها وإيقافها لخدمات وموارد الحوسبة السّحابيّة. وبالتالي هذا يؤدي إلى فقدان المال والمعلومات السرية والإنتاجيّة للمنظمات، أوللمستخدمين الذين يعتمدون على تخزين البيانات والخدمات والتطبيقات في الحوسبة السّحابيّة. وعملية إيقاف وكشف هجمات الدّودة في الحوسبة السّحابيّة أصبحت مسعاً وتحدياً صعباً لابدّ من دراستها والتركيز عليها، لذلك هذه الدراسة تقوم على تصنيف دودة الحوسبة السّحابيّة وبيان أساس خصائصها. بالإضافة إلى ذلك، فإنّ هذة الدراسة طوّرت تقنية إكتشاف الدّودة في الحوسبة السّحابيّة من خلال دّمج الخوارزميّة الجينيّة المحسّنة، وإقتراح تقنية الإستّجابة في الحوسبة السّحابيّة وبيان مستويات التّهديد. تمّ تقييم التقنيّة المقترحة؛ للكشّف والإستّجابة عن الدّودة في الحوسبة السّحابيّة إستناداً إلى معدلات الدقة. وتكمن قوة وإبداع التقنيّة المقترحة للكشّف والإستّجابة عن دودة الحوسبة السّحابيّة من خلال تصنيف دودة الحوسبة السّحابيّة، ودمج الخوارزميّة الجينيّة المحسّنة، وقياس مستوى التّهديد الذي يؤثر على سرية ونزاهة البيانات ومدى توافر البنية التحتية للحوسبة السّحابيّة. وفيما يتعلق في الخوارزميّة الجينيّة المحسّنة تم إنشاء متغيرات جديدة لتحسين الخوارزميّة الجينيّة على أساس إختيار الحلول النسبية الأمثل، والطّفرة الجينيّة، والتزاوج الجيني، ووحدة تحكم التطور الجيني بهدف زيادة مستوى الدّقة لتقنيّة الكشّف عن الدّودة في الحوسبة السّحابية. تقوم الخوارزميّة الجينيّة على فكرة عمليّة الإنتقاء الطّبيعي وعلم الوراثّة التي تقلّد العمليّة البيولوجيّة، مما يساعد على تحسين حل المشاكل المقيّدة، والغير مقيّدة. وطريقة إنشاء ذريّة جديدة من الأبناء يتم إستنباطها من خلال المزاوجة المثلى بين الآباء والأمهات لإستخدامها في الكشّف عن هجمات الديدان الغير معروفة و المستقبلية للحوسبة السّحابيّة. وقياس مستوى التّهديد للديدان في الحوسبة السّحابيّة إعتماداً على المقاييس الأمنيّة إستناداً إلى مستوى تأثير السريّة والنّزاهة للبيانّات والمعلومات، بالإضّافة إلى مدى توافر البنيّة التحتيّة للحوسبة السحابيّة. وتم إجراء التجارب في بيئة عمل مختبرية تحت الرقابة لتحليل ومراقبة سلوك الديدان في الحوسبة السّحابيّة وكان ذلك إعتماداً على تقنية إكتشاف المعرفة باستخدام أدوات مفتوحة المصدر. وقد استخدمت هذه الدّراسة 1195 مجموعة من العينّات في التّجارب حيث تم تطبيق التّحليل الدّيناميكي والمقاييس الأمنيّة. بناءً على نتائج التّجارب؛ أنتجت التقنيّة الخوارزميّة الجينيّة المحسّنة مايعادل 99.749% من معدّل كشف الدّقة للدّودة في الحوسبة السّحابيّة، بالإضافة إلى نسبة 0.014% من معدل الإكتشاف الكاذب، وبالتالي فإنّ هذه النتيجة حسّنت على تقنيّة الخوارزميّة الجينيّة أولكس بنسبة 41.05%. هذه الدّراسة ساهمت في تقدّيم تقنيّة جديدّة لدّيها القدّرة على الكشّف والإستّجابة لهجمات الدّودة في الحوسبة السّحابيّة.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

Page

ABBREVIATIONS

API	Application Program Interface
CSA	Cloud Computing Usage
DDoS	Distributed Denial-of-Service
DM	Data Mining
DoS	Denial-of-Service
DST	Dempster-Shafer Theory
DT	Decision Tree
EC2	Elastic Compute Cloud
EGA	Enhanced Genetic Algorithm
FAT	File Allocation Table
FN	False Negative
FP	False Positive
FTA	Fault-Tree Analysis
GA	Genetic Algorithm
IaaS	Infrastructure as a Service
IBK	K-nearest Neighbours
IDPS	Intrusion Detection and Prevention Systems
IDS	Intrusion Detection System
IRC	Internet Relay Chat
KDD	Knowledge discover database
LAN	Local Area Network
MD5	Message-digest 5
OlexGA	Olex Genetic Algorithm

PaaS	Platform as a Service
PE	Portable Executable
SaaS	Software as a Service
SMB	Server Message Block
TCP	Transmission Control Protocol
TN	True Negative
TP	True Positive
UDP	User Datagram Protocol
VM	Virtual Machine
VMM	Virtual Machine Monitor