RISK CONCENTRATION FOR CONTEXT ASSESSMENT (RiCCA) OF SMS MESSAGES USING DANGER THEORY

Kamahazira binti Zainal

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

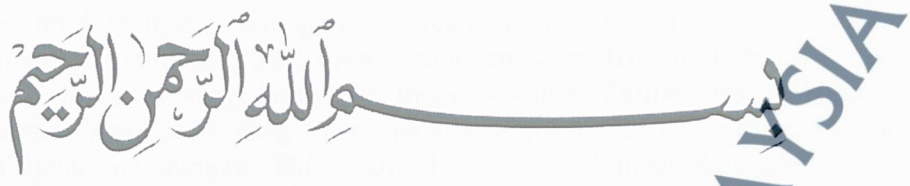
UNIVERSITI SAINS ISLAM MALAYSIA

August 2018

BIODATA OF AUTHOR

Kamahazira Zainal (4120261) was born on 29th March 1978. She previously was a student at Universiti Putra Malaysia (UPM) and obtained Bachelor of Engineering (Computer and Communication) in 2000. Then, she obtained her Master in Information Security from Universiti Teknologi Malaysia (UTM) in 2008. At this moment she is pursuing her Ph.D. from Universiti Sains Islam Malaysia (USIM), also specifically in information security and assurance.

ACKNOWLEDGEMENTS



Praise to Almighty Allah SWT for blessing me with good health, strength and determinations. The completion of this journey has not been possible without His will.

My greatest appreciation goes to my parents, especially my mom who always pray the best and keep the good thoughts. I also would like to express my deepest gratitude to my supervisor, Associate Professor Dr Mohd Zalisham bin Jali for his valuable guidance, trust, motivation and comments. The same wish goes to my co-supervisor, Ir. Dr Abu Bakar bin Hasan for his support and interest of my research. Without their continuous support, this thesis would not have been the same as presented here.

While working on this thesis, my research was supported by the Ministry of Higher Education of Malaysia and Research Management Centre of USIM under grant number USIM/FRGS/FST/32/50315. Thank you very much for the generous financial support.

To my beloved siblings and 'true' friends, thank you for your persistent support and understanding. I am very grateful to have all of you by my side especially during my difficulties and hard, tough time. To my teachers, this journey has begun because of you. For my darling nieces, Double MH (Marissa Humaira and Mardhiya El-Hannah binti Shahrul Azmizi), this wonderful journey is treasured cheerfully and accompanied with you two.

و ما توفيقي إلى بالله

And my success is from Allah.

ABSTRAK

Susulan daripada perkembangan teknologi telefon mudah alih dan Internet, *spam* merupakan satu isu serius yang telah timbul dengan banyak sekali. Rekod kerugian yang tidak pernah berakhir disebabkan oleh *spam* ini telah menggerakkan kajian ini. Kajian ini bertujuan untuk menjadi sebahagian pelengkap kepada penyelesaian sedia ada dan membantu pengguna mengenal pasti mesej *spam* yang berpotensi berbahaya. Tumpuan kajian ini adalah mesej *spam* berbahasa Inggeris dalam format teks Perkhidmatan Pesanan Ringkas atau SMS yang kebiasaannya tanpa disedari mengandungi unsur penipuan. Dengan kandungan SMS yang begitu meyakinkan dan keadaan para pengguna yang mudah tertipu dan tidak berwaspada merupakan salah satu faktor yang menyumbang kepada pelbagai kes berbentuk penipuan dan kecurian identiti. Kajian ini telah mengaplikasikan salah satu daripada teori *Artificial Immune Systems* (AIS), iaitu *Danger Theory*. Peranan *dendritic cells* yang berupaya untuk mengesan risiko kerosakan yang disebabkan oleh bahan berbahaya dalam tubuh manusia telah menjadi idea bagi teori ini. *Danger Theory* ini digunakan bagi menafsirkan kandungan mesej *spam* untuk mengesan sebarang kemungkinan risiko yang disebabkan oleh kandungan *spam* dengan mengilustrasikan elemen biologi dari teori ini di dalam konteks isu *spam*. Kombinasi teori biologi ini dengan teknik lain seperti penilaian risiko dan perlombongan teks telah menghasilkan model, *Risk Concentration for Context Assessment* atau RiCCA. RiCCA ini dibangunkan daripada algoritma *Danger Theory* iaitu *Dendritic Cell Algorithm* (DCA) dan *Deterministic Dendritic Cell Algorithm* (dDCA). Melalui beberapa siri ujian dengan menggunakan gabungan set data daripada *UCI Machine Learning* dan mesej *spam* yang dikumpulkan sendiri, telah menunjukkan RiCCA sebagai alat yang boleh mengukur tahap risiko *spam*. Kedua-dua algoritma ini berjaya menunjukkan bahawa *Danger Theory* mampu diaplikasikan dalam pengukuran risiko dengan kadar ketepatan adalah 90% dengan prestasi dDCA adalah lebih baik daripada DCA, kadar ketepatan 94.16%. Prototaip yang dibangunkan dapat dipertingkatkan lagi sebagai salah satu aplikasi mudah alih. Di samping itu, kajian ini boleh dilaksanakan dengan lebih lanjut untuk saiz mesej yang lebih besar (selain daripada SMS yang hanya terhad kepada 160 aksara) dan juga turut boleh diuji dalam bahasa selain daripada bahasa Inggeris.

ABSTRACT

Unified with the booming of mobile and Internet technology, spam is one of the serious issues that have been emerged tremendously. The never ending records of loss that caused by spam have initiated and motivated this study. With the purpose to produce a complementary solution for the current safeguards, this study aims to assist users in identifying potential harmful messages. The focus of the study is deploying an English text spam in Short Messages Services (SMS) format that usually consists of fraud intention in disguise. Due to the convincing but deceptive contents, users easily get enticed and their lack of awareness about implicit spam`s impact loss is one of the most vulnerable factors that lead to numerous cases of fraud, scam and identity theft. This study applied one of the prominent theories from Artificial Immune Systems (AIS), Danger Theory. The behaviour of dendritic cells to sniff danger that caused by a harmful substance in the human body has become the fundamental idea of this theory. Danger Theory is employed in deciphering the risky content of spam message via mapping its' biological properties in spam environment. Consolidation of this theory with other procedures such as risk assessment and text mining has produced a model namely as Risk Concentration for Context Assessment or RiCCA. This RiCCA prototype is developed from Danger Theory algorithms that is Dendritic Cell Algorithm (DCA) and Deterministic Dendritic Cell Algorithm (dDCA). Through a series of simulation with the deployment of merged dataset from UCI Machine Learning and self-collected spam messages, has indicated RiCCA as a reliable medium for measuring the risk concentration. Both algorithms demonstrated that Danger Theory is feasible in measuring risk of SMS content with more than 90% of accuracy rate, which dDCA outperformed DCA with distinctive result of 94.16% accuracy rate. For future work, the prototype can be further enhanced as one of the mobile application. Moreover, this study can be further applied for a larger size of message context (instead of SMS that is limited to 160 characters) and also tested in other languages.

AL-MULAKHKHAS الملخص

بعد تطوير تكنولوجيا الهاتف المحمول والإنترنت ، تعتبر الرسائل الاقتحامية مشكلة خطيرة نشأت إلى حد كبير إن السجل غير المنته من الخسارة بسبب الرسائل الاقتحامية قد نقل هذه الدراسة. تهدف الدراسة إلى أن تكون جزءًا مكملاً للحلول الحالية وتساعد المستخدمين على تحديد رسائل البريد العشوائي الخطيرة. التركيز على هذه الدراسة هو رسالة بريد إلكتروني غير مرغوبة باللغة الإنجليزية في تنسيق النص لخدمة الرسائل القصيرة أو وحالة SMS الرسائل النصية القصيرة التي عادة ما تحتوي على عناصر من الاحتيال. مع ما يسمى محتوى .المستخدمين الساذجين والمخفضين هي واحدة من العوامل التي تسهم في مختلف حالات الغش وسرقة الهوية نظرية الخطر. لقد أصبح دور الخلايا ، (AIS) طبقت هذه الدراسة واحدة من نظريات نظم المناعة الاصطناعية .التغصنية القادرة على اكتشاف خطر الأضرار الناجمة عن المواد الضارة في جسم الإنسان فكرة لهذه النظرية يتم استخدام نظرية الخطر هذه لتفسير محتويات رسائل البريد العشوائي للكشف عن أي خطر محتمل ناجم عن .محتوى البريد العشوائي من خلال توضيح العناصر البيولوجية لهذه النظرية في سياق قضايا البريد العشوائي ، هذا المزيج من النظريات البيولوجية مع تقنيات أخرى مثل تقييم المخاطر وتعدين النص قد أسفر عن نموذج من خوارزمية نظرية الخطر وهي خوارزمية RiCCA تم تطوير هذا .RiCCA تركيز المخاطر لتقييم السياق أو من خلال سلسلة من الاختبارات .(dDCA) وخوارزمية الخلية الشجيرية الديتريتيكية (DCA) الخلية التغصنية ورسائل الرسائل غير المرغوب UCI Machine Learning باستخدام مجموعة من مجموعات البيانات من .كأداة يمكن أن تقيس مستوى مخاطر الرسائل غير المرغوب فيها RiCCA فيها المجمعة نفسها ، فقد أظهرت وقد أظهرت كلتا الخوارزميتين أنه يمكن تطبيق نظرية الخطر في قياس المخاطر مع دقة معدل 90٪ مع أداء ومعدل دقة 94.16٪. يمكن زيادة تحسين النماذج الأولية كواحد من تطبيقات ، DCA أفضل من dDCA الجوال. بالإضافة إلى ذلك، يمكن تنفيذ هذه الدراسة بشكل أكبر لأحجام الرسائل الكبيرة (إلى جانب الرسائل .النصية القصيرة التي لا تتعدى 160 حرفًا) ويمكن اختبارها بلغات أخرى غير الإنجليزية

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ALGORITHMS

Algorithms **Page**

LIST OF APPENDICES

LIST OF PUBLICATIONS

1. **Zainal, K.,** Sulaiman, N. F., & Jali, M. Z. 2015. "An Analysis of Various Algorithms For Text Spam Classification and Clustering Using RapidMiner and Weka". *International Journal of Computer Science and Information Security (IJCSIS)*, *13*(3), 66–74. Retrieved from http://sites.google.com/site/ijcsis/

2. **Zainal, K.,** & Jali, M. Z. 2015. "A Perception Model of Spam Risk Assessment Inspired by Danger Theory of Artificial Immune Systems". In *International Conference on Computer Science and Computational Intelligence (ICCSCI)* (Vol. 59, pp. 152–161). Elsevier Masson SAS. http://doi.org/10.1016/j.procs.2015.07.530

3. **Zainal, K.,** & Jali, M. Z. 2016. "A Review of Feature Extraction Optimization in SMS spam Messages Classification". In *International Conference on Soft Computing in Data Science (SCDS)* (Vol. 652, pp. 158–170). http://doi.org/10.1007/978-981-10-2777-2_14

4. **Zainal, K.,** & Jali, M. Z. 2017a. "The Design and Development of Spam Risk Assessment Prototype: In Silico of Danger Theory Variants". *International Journal of Advanced Computer Science and Applications (IJACSA)*, *8*(4), 401–410. Retrieved from www.ijacsa.thesai.org

5. **Zainal, K.,** & Jali, M. Z. 2017b. "The Significant Effect of Feature Selection Methods in Spam Risk Assessment using Dendritic Cell Algorithm". In *International Conference on Information and Communication Technology (ICoICT 2017)* (pp. 277–284).

6. **Zainal, K.,** Jali, M. Z., & Hasan, A. B. 2018. "Comparative Analysis of Danger Theory Variants in Measuring Risk Level for Text Spam Messages." In *International Symposium on Data Mining Applications (SDMA)* (pp. 133–152). http://doi.org/10.1007/978-3-319-78753-4_11

7. **Zainal, K.,** & Jali, M. Z. 2018. "An Immunological-based Simulation: A Case Study of Risk Concentration for Mobile Spam Context Assessment". In *International Journal of Advanced Science, Engineering and Information Technology (IJASEIT)* (pp. 732-742). Retrieved from http://dx.doi.org/10.18517/ijaseit.8.3.2719

GLOSSARY

Adaptive immunity	a secondary response when the pathogen encountered for the second time which the concentration of relevant antibodies increases rapidly. Also known as specific response
Anomaly threshold (t_m for DCA; T_k for dDCA)	the signature of malicious in numerical value that separates normal and abnormal antigen. The measured output signal that is greater than this value is considered as malignant substance
Antigen	specific molecules present on surface of pathogen and recognized by immune system. The term antigen is an abbreviation from the word antibody generator
Apoptotic	signal alert released by normal cell death (semi-mature), indicate safe context
B-cells	develop in the bone marrow and involve in humoral immunity
Co-Stimulation (CSM)	a signal that act as co-stimulator in T-cells and B-cells activation
CD 4-Helper	commander of the immune response and it detects infection and sounds the alarm for initiating both T-cell and B-cell responses
CD 8-Cytotoxic	detects and kills infected or damaged body cells
Cell mediated immunity	involves mostly T-cells and responds to any cell that displays aberrant MHC markers, including cells invaded by pathogens, tumour cells, or transplanted cells
Context Assessment	measurement of cumulative output signals and the greater of semi-mature or mature of these signals become the cell context for antigen assessment. This cell context is used to label all antigen collected by the DCs with the derived context value in between 0 to 1
Danger Model	a theory of how the immune system works. It is based on the idea that the immune system does not distinguish between self and non-self, but rather between things that might cause damage and things that will not
Danger signal	signify abnormal behaviour which is indicates of an anomaly that is less than PAMPs

Danger Theory (DT)	takes care of 'non-self but harmless' and of 'self but harmful' invaders into human body systems. The central idea is that the immune system does not respond to non-self but to danger
Dendritic Cell (DC)	responsible to digest antigen material and forward it to the T-cells of the immune system. Also acting as messengers between the innate and adaptive immune systems
Ham	a valid or non-spam substance (such as SMS message)
Humoral immunity	involves B-cells or antibodies that recognize antigens or pathogens that are circulating in the lymph or blood
Immature DC (iDC)	Initial state of dendritic cell is immature and at this stage the cell is collecting debris and exposing to some signals will determine its next stage, either mature or semi-mature. Also known as naïve DC.
Immunity	refers to the situation in which an organism can defy a contagious illness
Inflammation	the signal amplify the effects of the other categories of signals; PAMPs, Danger and Safe signals
Innate immunity	primary response, occurs when any type of pathogen appears in the body for the first time. Also known as non-specific response
In Silico	refers to characterize biological experiments carried out entirely in a computer or via computer simulation
In Vitro	refers to the technique of performing a given procedure in a controlled environment outside of a living organism (within the glass)
In Vivo	refers to examination using a whole, living organism as opposed to a partial or dead organism (within the living)
Libtissue	prototype software system for constructing second generation AIS and applying them to real-world problems especially in intrusion detection study
Lightweight Intrusion detection SYStem (LISYS)	other prototype of AIS version for computational purposes, besides libtissue
Mature DC (mDC)	The immature dendritic cell that exposed to a greater quantity of either PAMPs or Danger signals than Safe

	signals will be transformed to mature state. At this stage, dendritic cells have the ability to present antigen and activate T-cells
Mature Context Antigen Value (MCAV)	determine the intensity or degree of the detected danger with the mean value of context per antigen type
Major Histocompatibility Complex (MHC)	a set of cell surface proteins essential for the acquired immune system to recognize foreign molecules
Necrotic	signal alert released by abnormal cell death (mature), indicate danger context
Pathogens Associated Molecular Patterns signal (PAMPs)	signify abnormal behaviour which highly indicates of an anomaly
Pathogen	an organism that can cause disease
Risk	is a measure of the extent to which an entity is threatened by a potential circumstance or event
Risk Concentration	translated as the severity of a spam text message. In this study, the malignant level referring to the risk density based on three (3) levels of categorical data (high, medium, low level of risk) that comes together with a numerical value (in between 0 to 1). The closer the calculated value to 1, the higher the potential risk is anticipated
Safe signal	signify a safe or normal context
Semi-mature DC (smDC)	The immature dendritic cell that exposed to a greater quantity of Safe signals will be transformed to semi-mature state. At this stage, dendritic cells do not have the ability activate T-cells but still able to present antigen
Spam	irrelevant or unsolicited messages, commonly sent over the Internet, typically to a large number of users, for the purpose such as advertising, phishing, and spreading malware.
T-cells	develop in the thymus and in charge for response in cell mediated immunity

LIST OF ABBREVIATIONS

Acc	Accuracy
AIS	Artificial Immune Systems
APC	Antigen Presenting Cell
A2P	Application to Person
BIS	Biological Immune Systems
CIA	Confidentiality, Integrity, Availability
CISO	Chief Information Security Officer
CSM	Co-Stimulation
CHI^2	Chi-square
DAD	Disclosure, Alteration, Denial
DC / DCs	Dendritic Cell / Dendritic Cells
DCA	Dendritic Cell Algorithm
dDCA	Deterministic Dendritic Cell Algorithm
DT	Danger Theory
FEMA	Federal Emergency Management Agency
FN	False Negative
FP	False Positive
GSMA	Groupe Speciale Mobile Association or GSM Association
iDC	Immature Dendritic Cell / Naïve DC
IG Ratio	Information Gain Ratio
IGF	Internet Government Forum
ISO	International Organization for Standardization
IT	Information Technology
LISYS	Lightweight Intrusion detection SYStem
MCAV	Mature Context Antigen Value
MCMC	Malaysian Communications and Multimedia Commission or Suruhanjaya Komunikasi dan Multimedia Malaysia (SKMM) in Malay
mDC	Mature Dendritic Cell
MHC	Major Histocompatibility Complex
NB	Naïve Bayesian
NIST	National Institute of Standards and Technology
PAMPs	Pathogens Associated Molecular Patterns
PDCA	Plan, Do, Check, Action
RiCCA	Risk Concentration for Context Assessment
RM	Risk Management
S	Risk Scale
smDC	Semi-mature Dendritic Cell
SMS	Short Messaging Services
S_p	Spam probability
SPIM	SPam over Instant Messaging
SVM	Support Vector Machine
TF	Term Frequency
TLR	Toll-Like Receptor
T_k	Anomaly Threshold (dDCA)
t_m / t(m)	Anomaly Threshold (DCA)

TN	True Negative
TP	True Positive
URL	Uniform Resource Locator
VoIP	Voice over Internet Protocol
WM	Weight Matrix
Wi-Fi	Wireless Fidelity