A FRAMEWORK FOR CLASSIFICATION SOFTWARE SECURITY USING COMMON VULNERABILITIES AND EXPOSURES

NOR HAFEIZAH BINTI HASSAN

DOCTOR OF PHILOSOPHY

2018

Faculty of Information and Communication Technology

A FRAMEWORK FOR CLASSIFICATION SOFTWARE SECURITY USING COMMON VULNERABILITIES AND EXPOSURES

Nor Hafeizah binti Hassan

Doctor of Philosophy

2018

A FRAMEWORK FOR CLASSIFICATION SOFTWARE SECURITY USING COMMON VULNERABILITIES AND EXPOSURES

NOR HAFEIZAH BINTI HASSAN

A thesis submitted
in fulfillment of the requirements for the degree of
Doctor of Philosophy

Faculty of Information and Communication Technology

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2018

DECLARATION

I declare that this thesis entitle “A Framework For Classification Software Security Using Common Vulnerabilities And Exposures ” is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature : ....

Name : ....

Date : ....

APPROVAL

I hereby declare that I have read this thesis and in my opinion this thesis is sufficient in terms of scope and quality for the award of Doctor of Philosophy.

Signature : ....

Supervisor Name : ....

Date : ....

DEDICATION

Hassan Aman

and

Maimun Yusof

ABSTRACT

The main research aim is to investigate what information is necessary to make a formal vulnerability pattern representation. This is done through the usage of formal Backus-Naur-Form syntax for the execution and presented with newly created vulnerability flow diagram. Some future works were also proposed to further enhance the elements in the secured software process framework. This thesis focuses on the research and development of the design, formalization and translation of the vulnerability classification pattern through a framework using common vulnerabilities and exposures data. To achieve this aim, the following work was carried out. First step is to create and conceptualized necessary meta-process. Second step is to specify the relationship between the classifiers and vulnerability classification patterns. This inclusive of the investigation of vulnerability classification objectives, processes, classifiers and focus domains among prominent framework. Final step is to construct the framework by establishing the formal presentation of the vulnerability classification algorithm. The validation process was conducted empirically using statistical method to assess the accuracy and consistency by using the precision and recall rate of the algorithm on five data sets each with 500 samples. The findings show a significant result with precision's error rate or **p** value is between 0.01 and 0.02 with error rate for recall's error rate is between 0.02 and 0.04. Another validation was conducted to verify the correctness of the classification by using expert opinions, and the results showed that the ambiguity of several cases were subdue. Formal-based classification framework with notation may increase accuracy and visualization compared with hierarchy-tree only, but the conclusion remains tentative because of methodological limitation in the studies.

ABSTRAK

Tujuan utama penyelidikan ini adalah untuk menyiasat perincian yang diperlukan untuk membuat perwakilan formal corak kerentanan. Ini dilakukan melalui penggunaan sintaks Backus-Naur-Form untuk pelaksanaan dan diwasilahkan dengan pengenalan kepada rajah aliran rentan yang baru. Beberapa titipan kerja untuk masa depan juga dicadangkan untuk menambahbaik elemen-elemen dalam rangka kerja perisian jamin-selamat. Tesis ini memberi tumpuan kepada penyelidikan dan pembangunan reka bentuk, formalisasi dan terjemahan corak klasifikasi kerentanan melalui rangka kerja menggunakan data kerentanan umum dan kededahan lazim. Untuk mencapai matlamat ini, kerja-kerja berikut telah dijalankan. Langkah pertama adalah mewujudkan dan memberi konsep kepada meta-proses. Langkah kedua ialah menentukan hubungan antara pengelas dan corak pengelas kerentanan. Ini termasuklah kenalpasti objektif klasifikasi kerentanan, proses, klasifikasi dan fokus domain di antara rangka kerja-rangka kerja yang ada. Langkah terakhir ialah membina rangka kerja dengan menghasilkan paparan algoritma klasifikasi kerentanan formal. Proses pengesahan dijalankan secara empirikal menggunakan kaedah statistik untuk menilai ketepatan dan ketekalan algoritma berdasarkan pada kadar ketepatan dan panggil-balik ke atas lima set data, setiap satunya dengan 500 sampel. Hasil penemuan menunjukkan dapatan yang signifikan dengan kadar ralat ketepatan atau nilai ***p*** *adalah antara 0.01 dan 0.02 dan kadar ralat untuk kadar ralat panggil-balik adalah antara 0.02 dan 0.04. Satu lagi pengesahan telah dijalankan untuk menentusahkan jenis klasifikasi dengan menggunakan pendapat pakar, dan hasilnya menunjukkan bahawa ketidaktentuan beberapa kes telah dikurangkan. Justeru, rangka klasifikasi berasaskan formal dengan notasi boleh meningkatkan ketepatan dan visualisasi berbanding dengan secara hiraki sahaja, tetapi kesimpulannya adalah tentatif kerana batasan metodologi dalam kajian.*

ACKNOWLEDGEMENTS

Alhamdulillah. I would like to gratefully acknowledge the Ministry of Higher Education Malaysia for sponsoring this research. Also, to the Center for Graduate Studies for their commitment in assisting this work.

The success of this research was made possible with the encouragement, motivation, assistance by many individuals. First and foremost is Professor Datuk TS. Dr. Shahrin Sahib, who made my dream to pursue and accomplish the work on Secured Software field comes true with his continual study guidance despite his challenging busy schedule. My thanks to Professor Dr. Burairah Hussin for his determination in encouraging the publication and my deepest gratitude to colleagues of Faculty of Information Communication and Technology.

This thesis accomplishment was also endured by my dearest, Azman Awang Teh, Irfan, Anis, Alia, Imran, Amni and Ilyas.

Such a challenging writing companion, yet enjoyable!

TABLE OF CONTENTS

PAGE

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

BNF	-	Backus-Naur-Form
CFG	-	Context-free grammar
CVE	-	Common Vulnerability and Exposures
DSMarker	-	Domain Specific Marker
DSSchema	-	Domain Specific Schema
DSWordlist	-	Domain Specific Wordlist
NVD	-	National Vulnerability Database
OSVDB	-	Open Source Vulnerability Database
OWASP	-	Open Web Application Security Project
PA	-	Protection Analysis
RISOS	-	Research Into Secure Operating Systems
VulClaF	-	Vulnerability Classification Framework
VulClaP	-	Vulnerability Classification Pattern

LIST OF PUBLICATIONS

1. Nor Hafeizah Hassan, Nazrulazhar Bahaman, Burairah Hussin, Shahrin Sahib, (2018), Enhancing the Secured Software Framework using Vulnerability Patterns and Flow Diagrams, International Journal of Advanced Computer Science and Applications (IJACSA).

2. Nor Hafeizah Hassan, Shahrin Sahib, (2015), Assessing The Mapping Process Using Evaluation Criteria to Validate Case Study Results, Recent Advances in Computer Sciences, WSEAS.

3. Hassan, N.H. and Selamat,S.R. and Sahib,S., (2014), Establishing the Relationship in Vulnerability Classification for a Secure Software Testing, Atlantis-Press.

4. Hassan, N.H. and Selamat, S.R. and Sahib,S. and Hussin,B., (2014), Incorporating Evaluation Criteria in Meta-process of Classification to Increase the Acceptance Level, *13th International Conference on Applied Computer,* ACACOS 2014.

5. Hassan, N.H. and Selamat, S.R. and Sahib S. and Hussin, B., (2011), Towards Incorporation of Software Security Testing Framework in Software Development, Springer.

CHAPTER 1

INTRODUCTION

1.1 Background

In software application, it is observed that there are negative consequences when security is compromised. Security can be compromised when there is lack of understanding of the in hand situation. Various terms used for security and it's family, huge numbers of models and framework to refer to, had created confusions to the software practitioner to classify vulnerability that is accurate, consistence and correct.

It is observed that there is a challenge in forming a vulnerability classification scheme due to type of data used. For example, some vulnerability database like Common Vulnerabilities Exposures or CVE is very much using natural language structure but without proper English grammar as given in itś web page of (*Common Vulnerabilities and Exposures:The Standard for Information Security Vulnerability Names*, 2015). One way to extract the information is by using semantic analysis (Rebolloa et al., 2015). However, in security domain, some terms are used differently. For instance, the meaning of buffer overflow is to overwrite the adjacent memory by overrun buffer and is not simply means that buffer is more than full. Therefore, it is learned that the terms must be specified with related to predefined rules of information security. Another challenge was to formally translate the domain terms into a schema that can be translated to a workable engine to extract the vulnerability given a historical database as debated in (Shaikh and Sasikumar, 2015). Therefore, this study is to focus on this scenario.

1.2 Problem Statement

The current vulnerability classification suffered from multiple dimensions of classifiers. They are either too specific or too complex (Ruohonen et al., 2017; Tripathi and Singh, 2011). Or they were only for dedicated cases. This lead to disability to perform a detection or protection from next attack of vulnerability. The understanding of the taxonomy which also various, requires a formal classification that can be used for generic cases regardless of applications, mobiles, networks or other devices (Burger et al., 2014).

The above research statement is divided into three research problem (RP) and the summary of the above statements are illustrated in Table 1.1.

Table 1.1: Summary of research problems

RP	Research Problems (RP)
RP1	The current vulnerability classification use multiple dimensions of classifiers are the issues needed to be addressed (Carl et al., 1994; Aslam et al., 1996b; Tripathi and Singh, 2011; Du and Mathur, 1998)
RP2	Lack of generic and systematic process to describe the vulnerability classification process , which disable to be performed on other classes. (Jiwnani and Zelkowitz, 2002; Katrina et al., 2005; S et al., 2005; Eagle et al., 2006; Bazaz and Arthur, 2007)
RP3	There is an absent of formal application to translate the vulnerability classification into solutions. (Eagle et al., 2006; Bazaz and Arthur, 2007; Lowis and Accorsi, 2011; Leitner and Rinderle-Ma, 2014). Therefore, the vulnerability classification requires a comprehensive and viable process