ATTACK GRAPH CONSTRUCTION FOR ENHANCING INTRUSION PREDICTION BASED ON VULNERABILITIES METRICS

ZAID JASIM MOHAMMED AL-ARAJI

DOCTOR OF PHILOSOPHY

2023

Faculty of Information and Communications Technology

ATTACK GRAPH CONSTRUCTION FOR ENHANCING INTRUSION PREDICTION BASED ON VULNERABILITIES METRICS

Zaid Jasim Mohammed Al-Araji

Doctor of Philosophy

2023

ATTACK GRAPH CONSTRUCTION FOR ENHANCING INTRUSION PREDICTION BASED ON VULNERABILITIES METRICS

ZAID JASIM MOHAMMED AL-ARAJI

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

Faculty of Information and Communications Technology

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2023

DECLARATION

I declare that this thesis entitled “Attack Graph Construction for Enhancing Intrusion Prediction Based on Vulnerabilities Metrics“ 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.

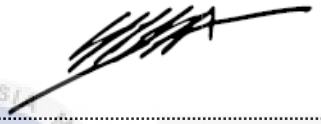
Signature : 

Name : Zaid Jasim Mohammed Al-Araji

Date : 12 March 2023

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 : Assoc. Prof. Dr. Sharifah Sakinah Syed Ahmad

Date : 12 March 2023

DEDICATION

To the one who gave me birth, to the reason for my existence in this life.

To the one who gives a taste to this life.

To the one that the tired of life filled his shoulders every morning to let me grow up.

To whom that Wrinkles filled his face and bleached his hair.

The reason for reaching this point, My Father, and Mother.

To those who make me laugh and support me always, my sister.

To the candle that removed the darkness of my way

To the wedges, that gave me strength, my teachers.

To the souls that support me near every edge, that hearts, those umbrellas that kept me away from the rains of despair, my beloveds, and friends.

I dedicate the hours of my effort to that manuscript on those papers of the past days.

ABSTRACT

The use of network technologies has increased in recent years. Although the network is beneficial for individuals to work and live in, it does have security challenges that should be rectified. One of these issues is cyberattacks. The attack surface for hackers is growing as more devices are linked to the internet. The next-generation cyber defence concentrating on predictive analysis seems more proactive than existing technologies based on intrusion detection. Recently, many approaches have been proposed to detect and predict attacks; one of these approaches is attack graphs. The main reason for designing the attack graph is to predict the attack as well as to predict the attack's next step in the network. The attack graph depicts the many paths an attacker may attempt to get around a security policy by leveraging interdependencies between disclosed vulnerabilities. The attack graph is categorized into three sections: generation, analysis, and use of attack graph. However, current attack graphs are suffering from a few issues. Scalability is the main issue the attack graph generation is facing. The reason for this issue is that the increase in the usage of devices connected to the network leads to increased vulnerabilities in the network, which leads to an increment in the complexity as well as generation time of the attack graph. However, the latest findings have employed the attack graph to forecast the next attack stage and manually locate the attack location for attack graph analysis. The attack graph is frequently employed in a few areas. Here, deriving security metrics is one component in which applying established security metrics might produce inaccurate findings. For this issue, this study proposes using intelligent agents to reduce the reachability time in calculating between the nodes and use the naïve approach prune algorithm to remove unnecessary edges, minimizing the attack graph's complexity. This study employs use Random Forest algorithm to identify and forecast attacks to dynamically locate the attack location in the network for attack graph analysis. The Weakest Path (WP), Mean Vulnerabilities on Path (MVoP), and Number of Vulnerabilities (NV) are three metrics introduced in this thesis. These metrics use network resources to determine the number of vulnerabilities and the network's weakest path. This work aims to generate a faster and less complexity attack graph and enhance the attack graph analysis to improve the detection and prediction of the attack, the attack's next step, and discover the weakest path that an attacker might use. For the results, the proposed attack graph performs better than the existing attack graph by using a naïve approach and a personal agent. The proposed attack graph reduced the generation time by 20% and the attack graph complexity. Besides, the RF algorithm produces encouraging results with an average accuracy rate of 97% in a different split of the CICIDS-2017 dataset and 94% using the CSE-CIC-IDS-2018 dataset. At the same time, vulnerabilities metrics provide better results and more understanding of the network. For future work, different pruning algorithms will be used to reduce the complexity, besides improving the attack prediction to increase the accuracy of determining the attack location.

PEMBINAAN GRAF SERANGAN BAGI PENAMBAHBAIKAN RAMALAN PENCEROBOHAN BERDASARKAN KELEMAHAN MATRIK

ABSTRAK

Penggunaan teknologi rangkaian telah meningkat dalam beberapa tahun kebelakangan ini. Walaupun rangkaian itu bermanfaat untuk individu untuk bekerja dan hidup, ia mempunyai cabaran keselamatan yang harus diperbetulkan. Salah satu isu ini ialah serangan siber. Permukaan serangan untuk penggodam semakin berkembang apabila lebih banyak peranti dipautkan ke internet. Pertahanan siber generasi akan datang yang menumpukan pada analisis ramalan nampaknya lebih proaktif daripada teknologi sedia ada berdasarkan pengesanan pencerobohan. Baru-baru ini, pelbagai pendekatan telah dicadangkan untuk mengesan dan meramalkan serangan; salah satu pendekatan ini ialah graf serangan. Sebab utama untuk mereka bentuk graf serangan adalah untuk meramalkan serangan serta langkah serangan seterusnya dalam rangkaian. Graf serangan menggambarkan pelbagai laluan yang mungkin cuba dilalui oleh penyerang untuk mengatasi polisi keselamatan dengan memanfaatkan kebergantung antara kelemahan yang didedahkan. Konsep graf serangan dikategorikan kepada tiga bahagian: penjanaan, analisis, dan penggunaan graf serangan. Walau bagaimanapun, graf serangan sedia ada mengalami beberapa isu. Kebolehskalaan ialah isu utama yang dihadapi oleh penjanaan graf serangan. Sebab bagi isu ini berpunca dari peningkatan dalam penggunaan peranti membawa kepada peningkatan keterdedahan dalam rangkaian, yang membawa kepada peningkatan dalam kerumitan serta masa penjanaan graf serangan. Penemuan terkini, bagaimanapun, telah menggunakan graf serangan untuk meramalkan peringkat serangan seterusnya dan memperuntukkan lokasi serangan secara manual untuk analisis graf serangan. Ia kerap digunakan di beberapa kawasan berkenaan dengan graf serangan. Di sini, memperoleh metrik-metrik keselamatan merupakan satu komponen di mana menggunakan metrik keselamatan yang ditetapkan mungkin menghasilkan penemuan yang tidak tepat. Untuk isu ini, kajian ini mencadangkan penggunaan ejen-ejen pintar untuk mengurangkan masa kebolehcapaian dalam mengira antara nod-nod dan menggunakan algoritma prun naïve *approach untuk mengeluarkan sisi tepi yang tidak perlu, yang mampu meminimumkan kerumitan graf serangan. Kajian ini menggunakan algoritma pembelajaran mesin untuk mengenal pasti dan meramalkan serangan untuk mengesan lokasi serangan dalam rangkaian secara dinamik bagi analisis graf serangan. Laluan Terlemah (WP), Min Keterdedahan pada Laluan (MVoP) dan Bilangan Keterdedahan (NV) merupakan tiga metrik yang diperkenalkan dalam kajian ini. Metrik-metrik ini menggunakan sumber rangkaian untuk menentukan bilangan keterdedahan dan laluan rangkaian yang paling lemah. Kerja ini bertujuan untuk menjana graf serangan yang lebih pantas dan kurang kerumitan serta mempertingkatkan analisis graf serangan untuk meningkatkan pengesanan dan ramalan serangan, langkah serangan seterusnya dan menemui laluan paling lemah yang mungkin digunakan oleh penyerang. Untuk keputusan, graf serangan yang dicadangkan berprestasi lebih baik daripada graf serangan sedia ada dengan menggunakan pendekatan naif dan ejen peribadi. Graf serangan yang dicadangkan mengurangkan masa penjanaan sebanyak 20% dan kerumitan graf serangan. Selain itu, algoritma RF menghasilkan hasil yang menggalakkan dengan kadar ketepatan purata 97% dalam pembahagian berbeza bagi dataset CICIDS-2017 dan 94% menggunakan dataset CSE-CIC-IDS-2018. Pada masa yang sama, metrik kelemahan memberikan hasil yang lebih baik dan*

lebih memahami rangkaian. Untuk kerja masa hadapan, algoritma pemangkasan berbeza akan digunakan untuk mengurangkan kerumitan, selain menambah baik ramalan serangan untuk meningkatkan ketepatan menentukan lokasi serangan.

ACKNOWLEDGEMENTS

In the Name of Allah, the Most Gracious, the Most Merciful

First and foremost, I would like to thank and praise Allah the Almighty, my Creator, my Sustainer, for everything I have received since the beginning of my life. Alhamdulillah, all praise and gratefulness are due to Allah. Thank you, Allah, for giving me good health and strength throughout my journey in completing this study.

I would like to extend my appreciation to the Universiti Teknikal Malaysia Melaka (UTeM) for providing the research platform. I also would like to offer my sincerest gratitude to my supervisor, Assoc. Prof. Dr. Sharifah Sakinah Syed Ahmad who has supported me and guided me with her patience, knowledge, and experience. I gratefully acknowledge her for her supervision, advice, and contribution. Also, to my co-supervisor, Ts. Dr. Raihana Syahirah Abdullah, who constantly supported my journey.

My special thank goes to my family members, especially my parent and sister, for their inseparable, inspiring prayers. It is also a pleasure to express my appreciation to my friends for supporting me and encouraging me to complete this project. Without their support and encouragement, I would not have been able to complete it.

Finally, I would like to thank everybody important who helped me to complete my project, as well as express my apology.

Thank you all.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACO	-	Ant Colony Optimization
ACO-MCGS	-	Ant Colony Optimization using Monte Carlo Graph Search
ACT	-	Attack Countermeasure Tree
ADT	-	Attack Defence Tree
AFT	-	Attack Fault Tree
AGFSM	-	Attack Graph Model Based on Finite State Machine
AHAG	-	Automatic Hybrid Attack Graph
AHE	-	Acceptable Head Edge
AI	-	Attack Impact
ANN	-	Artificial Neural Network
APV	-	Attack Path Variability
APVIS	-	Attack Path Variability with IP Shuffling
ARIMA	-	AutoRegressive Integrated Moving Average
ART	-	Attack Response Tree
AT	-	Attack Tree
BN	-	Bayesian Network
CIC	-	Canadian Establishment for Cybersecurity
CNN	-	Convolutional Neural Network
CPA	-	Coloured Petri Automata
CPS	-	Cyber-Physical System
CSP	-	Compromise Success probability
CVE	-	Common Vulnerabilities and Exposures
CVSS	-	Common Vulnerability Scoring System

CWE	-	Common Weakness Enumeration
CWSS	-	Common Weakness Scoring System
DAG	-	Defence Attack Graph
DBN	-	Deep Brief Network
DDoS	-	Distributed Denial-of-Service
DFW	-	Distributed Firewall
DNN	-	Deep Neural Network
DT	-	Defence Tree
EPGM	-	Extended Property Graph Model
FN	-	False Negative
FoS	-	Factor of Security
FP	-	False Positive
FSM	-	Finite State Machine
FT	-	Fault Tree
GAN	-	Generative Adversal Network
GCON	-	Graph Constrain Specification Language
GED	-	Graph Edit Distance
GM	-	Grey Model
GSM	-	Graph Security Model
HAG	-	Hybrid Attack Graph
HARM	-	Hierarchical Attack Representation Model
HAV	-	Host Address Variability
HIDPS	-	Host-Based Intrusion Detection and Prevention System
HTTP	-	Hypertext Transfer Protocol

ICS	-	Industrial Control System
IDPSs	-	Intrusion Detection and Prevention Systems
IDS	-	Intrusion Detection System
IPS	-	Intrusion Prevention System
KNN	-	K- Nearest Neighbour
LR	-	Logistic Regression
MAPL	-	Mean of Path Lengths
MCS	-	Maximum Common Subgraph
MIT	-	Matching Index Table
MoPL	-	Mode of Path Lengths
MPL	-	Mean of Path Length
MTD	-	Moving Target Defense
MTD	-	Moving Target Defense
MulVAL	-	Multi-host, Multi-stage Vulnerability Analysis Language
MV-HARM	-	Maritime Vessel-Hierarchical Attack Representation Model
MVoP	-	Mean Vulnerabilities on Path
NAP	-	Number of Attack Paths
NBA	-	Network Behaviour Analysis
NetSPA	-	Network Security Planning Architecture
NFS	-	Network File System
NIDPS	-	Network-Based Intrusion Detection and Prevention System
NIST	-	National Institute of Standards and Technology
NMPL	-	Normalised Mean of Attack Path Lengths
NSS	-	Number of Severe Systems

NV	-	Number of Vulnerabilities
NVAG	-	Network Vulnerability Attack Graph
NVD	-	National Vulnerability Database
NVM	-	Normalized Vulnerability Metric
OWA	-	Ordered Weighted Averaging
OWAT	-	Ordered Weighted Averaging Tree
OWL	-	Web Ontology Language
P-BEST	-	Production-Based Expert System Toolest
PFoS	-	Probabilistic Factor of Security
PSS	-	Percentage of Severe Systems
PT	-	Protection Tree
RBN	-	Restricted Boltzmann Machine
RDF	-	Resource Description Framework
RF	-	Random Forest
RNN	-	Recurrent Neural Network
RUSSEL	-	Rule-Based Sequence Evaluation Language
S3	-	Scalable Security State
SAP	-	Shortest Attack Path
SAPV	-	Shortest Attack Path Variability
SAPVIS	-	Shortest Attack Path Variability with IP Shuffling
SDA	-	Software Diversity-Based Adaptation
SDN	-	Software Defined Network
SDPL	-	Standard Deviation of Paths Lengths
SQL	-	Structured Query Language

SSP	-	Scan Success Probability
SVM	-	Support Vector Machine
TAG	-	Topological Attack Graph
TCP/IP	-	Transmission Control Protocol/Internet Protocol
T-HARM	-	Temporal Hierarcal Attack Representational Model
TN	-	True Negative
TNH	-	Total Number of Network Hosts
TP	-	True Positive
TTPs	-	Tactics, Techniques, and Procedures
TVA	-	Topological Vul
UEFI	-	Unified Extensible Firmware Interface
VNAP	-	Variability of the Number of Attack Paths
WIDPS	-	Wireless Intrusion Detection and Prevention System
WP	-	Weakest Path

LIST OF PUBLICATIONS

The followings are the list of publications related to the work on this thesis:

1. Z. J. Al-araji, S. S. A. Syed, M. W. Al-salihi, H. A. Al-lamy, M. Ahmed, and W. Raad, 2019, "Network Traffic Classification for Attack Detection Using Big Data Tools : A Review," *Intelligent and Interactive Computing, Lecture Notes in Networks and Systems*, 67, pp. 355–363.

2. Z. J. Al-Araji, S. S. S. Ahmad, and R. S. Abdullah, 2021, "Comparison Study Between Attack Graph Path Based Metrics," *in Proceedings of the 3rd International Conference on Intelligent and Interactive Computing*, pp. 10-13.

3. Z. J. Al-Araji, S. S. S. Ahmad, R. S. Abdullah, A. A. Mutlag, H. A. A. Raheem, and S. R. H. Basri, 2021, "Attack graph reachability: concept, analysis, challenges and issues," *Network Security*, 2021(6), pp. 13–19.

4. Z. J. Al-Araji, S. S. S. Ahmad, and R. S. Abdullah, 2021, "Propose Vulnerability Metrics to Measure Network Secure using Attack Graph," *International Journal of Advanced Computer Science and Applications*, 12(5), pp. 51–58.

5. Z. J. Al-Araji, S. S. S. Ahmad, and R. S. Abdullah, 2022. "Attack Prediction to Enhance Attack Path Discovery Using Improved Attack Graph". *Karbala International Journal of Modern Science*, 8(3), pp. 313-329.