Modeling Cyber Attacks with Empirical Correlation

Author

Kamran Saeed

01-242171-009

Supervisor:

Dr. Muhammad Najam ul Islam

Co-Supervisor:

Dr. Mureed Hussain

This dissertation is submitted for the degree of

MS Computer Engineering

Department of Computer Engineering

Faculty of Engineering Sciences

Bahria University, Islamabad Campus, Pakistan

March, 2019

In the name of Allah, the Most Gracious, the Ever Merciful

Bahria University
Discovering Knowledge

MS-13

Thesis Completion Certificate

Student's Name: **Kamran Saeed** Registration No. **49954**
Programme of Study: **MS Computer Engineering**
Thesis Title: **"Modeling Cyber Attacks with Empirical Correlation"**

It is to certify that the above student's thesis has been completed to my satisfaction and, to my belief, its standard is appropriate for submission for Evaluation. I have also conducted plagiarism test of this thesis using HEC prescribed software and found similarity index at **08%** that is within the permissible limit set by the HEC for the MS/MPhil degree thesis. I have also found the thesis in a format recognized by the BU for the MS/MPhil thesis.

Principal Supervisor's Signature: _____

Date: _____ Name: **Dr. M. NAJAM UL ISLAM**

Co-Supervisor's Signature: _____

Date: _____ Name: **Dr. MUREED HUSSAIN**

Bahria University
Discovering Knowledge

MS-14A

Author's Declaration

I, __**Kamran Saeed**_____ hereby state that my MS thesis titled _**"Modeling Cyber Attacks with Empirical Correlation"**_____ is my own work and has not been submitted previously by me for taking any degree from this university_____**Bahria University**___or anywhere else in the country/world.

At any time if my statement is found to be incorrect even after my graduation, the university has the right to withdraw/cancel my MS degree.

Author's Signature:_____

Name of student: __KAMRAN SAEED__

Date: _____

Bahria University
Discovering Knowledge

MS-14B

Plagiarism Undertaking

I, solemnly declare that research work presented in the thesis titled "**Modeling Cyber Attacks with Empirical Correlation**" is solely my research work with no significant contribution from any other person. Small contribution / help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of the HEC and Bahria University towards plagiarism. Therefore I as an Author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS degree, the university reserves the right to withdraw / revoke my MS degree and that HEC and the University has the right to publish my name on the HEC / University website on which names of students are placed who submitted plagiarized thesis.

Student / Author's Sign: _____

Name of the Student: KAMRAN SAEED

Abstract

Cyber-attacks have been on the rise especially after the explosive widespread of social networking as it gives cyber criminals a way to break into other's computers and manipulate personal and sensitive data. Many different techniques have been used in the past to minimize the occurrences of cyber-attacks. These techniques focused primarily on attack modeling by analyzing the incoming traffic in order to look for both malicious activity and attacker objectives. This research proposes a solution that makes use of the attack tree modeling (ATM) along with the development of a correlation engine that predicts coordinated attacks carried out on network servers. The correlation engine uses network flow features i.e. control information about the transmitting content and correlates them based on the previously learned labeling to see if the content is malicious or not. The correlation engine can predict Distributed Denial of Service (DDOS) and Brute-force attacks. These attack categories have been separately modeled using the highest real-time traffic performance algorithm out of Support Vector Machine (SVM), Gaussian Naive Bayes (GNB) and Random Forest Regression (RFR) techniques. The correlation engine tests real-time data and along with the prediction of attacks, it also updates the stored labeling based on system administrator feedback. Once deployed, the correlation engine can be used in real-time on any network or server to continuously monitor and detect zero-day attacks that undermine the integrity of the network or its data.

Table of Contents

1 Introduction 1
1.1 Cyber Attacks 1
1.2 Types of cyber attacks 2
1.2.1 Cyber-attacks based on behavior 2
1.2.2 Cyber-attacks based on medium 3
1.2.3 Most common cyber-attacks 3
1.3 Problem statement 6
1.4 Motivation 7
1.5 Thesis Structure 8
1.6 Major Contributions 8
2 Literature Review 9
2.1 Approaches of correlation 15
2.1.1 Similarities of Alert Correlation 15
2.1.2 Prerequisites and Consequences of Attacks 16
2.1.3 Predefined Attack Scenario 16
2.1.4 Expert Systems and Data Mining 16
2.2 Literature Synthesis for correlation techniques 17
3 Proposed Methodology 22
3.1 Machine Learning Algorithms 23
3.1.1 Support Vector Machine 23
3.1.2 Gaussian Naïve Bayes 24
3.1.3 Random Forest Regression 24
4 Implementation 27
4.1 Attack Tree 27
4.1.1 Data Leakage 27
4.1.2 Data Modification 28

4.1.3 Data Theft 29
4.1.4 System attacks 29
4.2 Correlation Engine 31
4.3 Correlation Engine Architecture 32
4.3.1 Model Training 32
4.3.2 Real-time Attack Prediction 33
4.3.3 Self-learning Mode 34
4.4 Dataset Features 38
4.4.1 User Datagram Protocol (UDP) Flooding 38
4.4.2 Transmission Control Protocol (TCP) Flooding 39
4.4.3 Internet Control Message Protocol (ICMP) Flooding 40
4.4.4 Hypertext Transfer Protocol (HTTP) Flooding 40
4.5 Brute-force and Password Cracking Attack 41
5 Results and Analysis 43
5.1 Accuracy 43
5.2 Confusion Matrix 44
5.3 Choice of Learning Techniques 45
5.3.1 User Datagram Protocol (UDP) flooding 45
5.3.2 Transmission Control Protocol (TCP) flooding 47
5.3.3 Internet Control Message Protocol (ICMP) flooding 49
5.3.4 Hypertext Transfer Protocol (HTTP) flooding 51
5.3.5 Brute-force and Password Cracking Attacks 56
5.4 Real-time Normal Traffic Testing 59
5.4.1 Comparison of TCP Models 60
5.4.2 Comparison of UDP Model 61
5.4.3 Comparison of ICMP Model 62
5.4.4 Comparison of HTTP Model 63

5.5 Real-time Attack Testing 64
5.5.1 Comparison of TCP Model 65
5.5.2 Comparison of UDP Model 66
5.5.3 Comparison of ICMP Model 68
5.5.4 Comparison of HTTP Model 69
5.6 Overall Real-time Traffic Testing 71
5.6.1 Comparison of TCP Model 71
5.6.2 Comparison of UDP Model 72
5.6.3 Comparison of ICMP Model 73
5.6.4 Comparison of HTTP Model 74
5.7 Discussion 75
6 Conclusion 80
7 Future works 82
8 References 83