Multiple Simultaneous Threats Detection in Distributed Systems

By

Zafar Sultan

MSc Computer Science

School of Science and Technology
University of New England

A thesis submitted for the degree of

PhD in Computer Science

of University of New England

March, 2011

Declaration

I certify that the substance of this thesis has not already been submitted for any degree and is not currently being submitted for any other degree or qualification.

I certify that any help received in preparing this thesis, and all sources used have been acknowledged in this thesis.

Signature

Acknowledgments

I would like to thank Dr. Paul Kwan (principal supervisor) for his suggestion for an initial research direction and his valuable comments on how to structure this thesis. In fact, this thesis would have not been possible without his guidance and direction. Thanks to John Revington for proofreading and editorial advice.

I am thankful to many professionals who have helped and supported me through this doctoral study at the University of New England. First of all, I appreciate IBM colleagues for their excellent guidance and encouragement, their patience in discussing the research experiments, the methodology and results, and their original suggestions on some of the solutions during my PhD study. Their deep insights on the subject and on the architecture of the UNIX environments have been an important resource.

I would like to express gratitude to the IBM Architecture Team for their useful tips on the experimental layout and design of the multiple simultaneous threats detection system. I also thank the members of the IBM UNIX Delivery Team for discussing and reviewing the UNIX environments setup for the multiple simultaneous threats detection system. I would also like to thank the members of the IBM Network Team for reviewing the work and suggesting an optimum network setup for my multiple simultaneous threats detection system. I also extend thanks to the members of the IBM Windows Team for discussions of Windows integration with the UNIX environments that was the foundation of the

multiple simultaneous threats detection system. I also thank Dr Mark Evered, co-supervisor, for his valued comments during the literature review.

I owed a lot to the friends for their warm friendships and help in both professional and personal circumstances. I would like to thank the University of New England's administration staff. They have made my stay at Armidale an enjoyable experience during the Faculty of Arts and Science conferences in 2008 and 2009.

Finally, I would like to express thanks to my wife and children wholeheartedly for their endless love and support during this study. I thank all of my friends for encouraging me to complete this PhD thesis.

Publications

PhD Work published in Journal Papers

1) Sultan, Zafar and Paul W. Kwan, Generalized Evidential Processing in Multiple Simultaneous Threat Detection in UNIX, International Journal of Web Portals (IJWP), Volume 2, Issue 2, IGI Publishers, page(s) 51-67, 2010

2) Sultan, Zafar, Multiple Simultaneous Threat Detection in UNIX Environment. IJCSNS International Journal of Computer Science and Network Security, VOL.9 No.2, page(s) 65-75, February 2009

3) Sultan, Zafar, Survey and Research Directions on Intrusion Detection in UNIX, Environment. IJCSNS International Journal of Computer Science and Network Security, VOL.9 No.12, page(s) 69-74, December 2009

Abstract

This research examines a simultaneous threats detection system for distributed systems that uses a hybrid identification fusion model. This hybrid model is comprised of mathematical and statistical data fusion engines: Dempster-Shafer, Extended Dempster-Shafer, and Generalised Evidential Processing (GEP). The simultaneous threats detection system produced threat detection rates of 56% using Dempster-Shafer whilst Extended Dempster-Shafer and Generalised Evidential Processing (GEP) achieved 80% and 95% threat detection rate. Thus, the simultaneous threats detection system can improve threat detection rates by 39% (i.e. 95% - 56%) simply by adopting a more effective hybrid fusion model. In terms of efficiency and performance, the comparison of the three inference engines of the simultaneous threats detection system showed that Generalised Evidential Processing is a better data fusion model than Dempster-Shafer or Extended Dempster-Shafer.

In addition, the set cover packing technique was used as a middle-tier data fusion tool to determine the reduced size groups of the threat data. Set cover provided significant improvement and reduced the threat population from 2,272 to 295. This helped to minimise the complexity of evidential processing, and therefore reduced the cost and time taken to determine the combined probability mass of the multiple simultaneous threats detection system. This technique is particularly relevant to online and internet-dependent applications, including portals.

Key words:

Multiple Simultaneous Threats Detection; Distributed Systems; Intrusion Detection Systems; Bayesian Theory; Dempster-Shafer; Multisensor Data Fusion; Extended Dempster-Shafer; Set Cover; Set Packing; Generalised Evidential Processing; UNIX; Portals.

Table of Contents

DECLARATION i
ACKNOWLEDGMENTS ii
Publications iv
Abstract v
List of Figures x
List of Tables xi
1. INTRODUCTION 1
1.1 RESEARCH PROBLEMS 3
1.2 EXISTING SOLUTIONS 5
1.3 THE IMPORTANCE OF MULTIPLE SIMULTANEOUS THREATS DETECTION 7
1.4 APPROACH AND METHODOLOGY 9
1.5 SIGNIFICANCE AND INNOVATION 13
1.6 CONTRIBUTIONS OF THIS THESIS 15
1.7 ORGANISATION OF REMAINING CHAPTERS 16
2. LITERATURE REVIEW 17
2.1 PROBLEMS IN INTRUSION DETECTION 17
2.2 MULTISENSOR DATA FUSION 19
2.3 THE IMPORTANCE OF SIMULTANEOUS THREATS DETECTION 24
2.4 RELATED WORK 27
2.5 RESULTS AND BENEFITS 32
3. MULTIPLE SIMULTANEOUS THREATS DETECTION SYSTEM 33
3.1 TYPES OF THREATS 34
3.2 THREATS GENERATION UTILITIES 35
3.3 WHY ONLY FOUR THREATS 36
3.4 INTRUSIONS DETECTION SYSTEMS AS INDEPENDENT OBSERVERS 36
4. DATA FUSION PROCESS MODELS 38
4.1 ARCHITECTURE OF THE MULTISENSOR DATA FUSION PROCESS MODEL 38
4.2 BRIEF DESCRIPTION OF THE MULTISENSOR DATA FUSION PROCESS MODEL 39
4.3 SET COVER THEORY 39
4.3.1 Overview of Set Cover 39
4.3.2 Greedy Algorithm 41
4.3.4 Empirical Tests 43
4.3.5 Benefits of Set Cover as a Middle-tier Data Fusion Tool 45
4.4 THE DEMPSTER-SHAFER FUSION MODEL 47
4.4.1 Overview of the Dempster-Shafer Theory 47

4.4.2 Evidence to Proposition Assignments 48
4.4.3 Threats as Propositions in Intrusion Detection 49
4.4.4 Limitations of the Dempster-Shafer Theory 52
4.4.5 Fusion Without Considering Weights of Each Sensor 53
4.4.6 Dempster-Shafer Combined Probability Mass Functions 54
4.4.6.1. An Example of Two Threats 55
4.4.6.2. An Example of Three Threats 57
4.4.6.3. An Example of Four Threats 58
4.4.7 Limitation of Proposed Enhancements 60
4.5 EXTENDED DEMPSTER-SHAFER THEORY TO FUSE DATA 60
4.5.1 Overview of the Extended Dempster Shafer Theory 61
4.5.2 Evidence to Proposition Assignments 61
4.5.3 Fusion With Considering Weight of Each Sensor 62
4.5.3.1. Determining the Weights of Observations 64
4.5.3.2. Limitations in Calculating Weights 65
4.5.3.3. Note on Generalized Evidential Processing 65
4.5.4 Extended Dempster-Shafer Enhanced With Weights 66
4.5.4.1. An Example of Two Threats 68
4.5.4.2. An Example of Three Threats 69
4.5.4.3. An Example of Four Threats 72
4.6 DATA FUSION USING GENERALIZED EVIDENTIAL PROCESSING THEORY 74
4.6.1 Overview of the Generalized Evidential Processing Theory 74
4.6.2 Empirical Assessment 75

5. EXPERIMENTAL EVALUATION 84

5.1 EXPERIMENTAL SETUP 86
5.1.1 MSTDS Process Model of the Experiment 86
5.1.2 Context Diagram 87
5.1.2.1. Description of the Context Diagram 88
5.1.3 Multiple Simultaneous Threats Detection Process Model 89
5.2 EXPERIMENTAL RESULTS 90
5.2.1 Based on Dempster-Shafer Fusion Model 90
5.2.2 Based on Extended Dempster-Shafer Fusion Model 91
5.2.3 Based on Generalised Evidential Processing Fusion Model 92
5.3 VERIFICATION OF MSTDS USING PUBLIC DOMAIN DATA SETS 93
5.3.1 Data Fusion Process 94
5.3.2 An Example of Threat detection with two Threats 96
5.4 COMPARISONS WITH RELATED WORKS 102
5.4.1 Description of the Related Works 102
5.4.2 Results 103
5.4.2.1. Data Fusion Model Approaches 103
5.4.2.2. Comparisons of the Performance of the Inference Theory 104
5.4.2.3. Decision Level Techniques 105
5.5 DISCUSSION 108
5.5.1 Limitations of the Proposed MSTD Model 113

6. CONCLUSIONS 116

6.1 FUTURE WORK 118
BIBLIOGRAPHY 121

List of Figures

Fig 3.1: Threat nodes used in experiments of the multiple simultaneous threats detection system 34

Fig 4.1: Architecture of the Multiple Simultaneous Threats Detection System. 38

Fig 4.2: Z (Set Cover) and Q (sets with minimum cost) of the Universal Set ... 44

Fig 5.1: Multisensor Data Fusion Process Model of the Experiment..... 86

Fig 5.2: Infrastructure of the MSTDS Environment..... 87

Fig 5.3: Experimental Environment for the MSTDS 89

Fig 5.4: Performance Comparisons between Bayesian and Dempster-Shafer . 104

Fig 5.5: Comparison of MSTDS Decision Level Techniques..... 107

Fig 5.6: Results based on Dempster Shafer (unweighted) and Extended Dempster Shafer (weighted)..... 108

Fig 5.7: Effectiveness of the Multiple Simultaneous Threats Detection System 110

Fig 5.8: Performance of the Multiple Simultaneous Threats Detection Systems 111

Fig 5.9: Performance Trend of the Multiple Simultaneous Threats Detection System 112

Fig 5.10: Overall Performance comparisons of the IDS methods..... 113

List of Tables

Table: Definitions & Acronyms xii

Table 4.1: Set Cover reduces the sizes of the subsets of threat data .. 42

Table 4.2: Summary of the probability mass calculation by MSTDS ... 83

Table 5.1: Threat Results based on Dempster-Shafer Theory of Inference 91

Table 5.2: Threat Results based on Extended Dempster Shafer Theory of Inference 92

Table 5.3: Threat Results based on Generalised Evidential Processing 93

Table 5.4: Set Cover reduces the size of the data of Public Domain Data Sets 94

Table 5.5: Comparisons of the Combined Probability Masses of MSTDS Experimental Data and Public Domain Data sets 101

Table 5.6: Threat Detection Rate using Dempster Shafer on Public Domain Data Sets 101

Table 5.7: Threat Detection Rate using Extended Dempster Shafer on Public Domain Data Sets 101

Table 5.8: Threat Detection Rate using GEP on Public Domain Data Sets 102

Table: Definitions & Acronyms

Abbreviation	Definition
ACL	Access Control List
Bel	Belief
bpa	Basic probabilty assignment or m
Buffer Overflow	A buffer overflow occurs when a program or process tries to store more data in a buffer (temporary data storage area) than the buffer was designed for
CPU	Central Processing Unit
CSA	Cisco Security Agent
DA	Detected Alerts
DEV	Development Computing Environment
DNS	Domain Name Services
DoS	Denial Of Service
DS	Dempster-Shafer
FC	Fibre Channel - Gigabit speed network technology primarily used for storage networking
FPR	False Positive Rates
FTP	File Transfer Protocol
GEP	Generalised Evidential Processing
IBM	International Business Machines Corporation
ICT	Information and Communications Technology
IDS	Intrusion Detection System
IT	Information Technology
MARS	Cisco Security 'Monitoring, Analysis, and Response System' (Security monitoring for threat identification, mitigation and compliance secures network devices and host applications)
MaxEnt	maximum entropy (a postulate about a universal feature of any probability assignment on a given set of propositions)
MITMA	man-in-the-middle attack
MMSE	minimum mean square error (a statistical procedure)
MSTDM	Multiple Simultaneous Threat Detection Model
MSTDS	Multiple Simultaneous Threat Detection System
OA	Observed Alerts
OS	Operating System
OSI	Open Systems Interconnection
OT	Observed Threat
Pl	Plausibility
PPP	Point-to-Point Protocol (Internet)
QoS	Quality of Service
RA	Real Alerts
Set Cover	The set covering problem is a classical question in computer science and complexity theory that has led to the development of fundamental techniques for the entire field of approximation algorithms
Sniffers	Scotland & Northern Ireland Forum for Environmental Research (a sniffer is a program that monitors and analyses network traffic, detecting bottlenecks and problems)
SNMP	Simple Network Management Protocol
Snoop	Packet sniffer
TCP/IP	Transmission Control Protocol/Internet Protocol
Trojan Horse	Trojans allow a hacker remote access to a target computer system
UNE	University Of New England, Armidale, NSW, Australia

UNIX	Uniplexed Information and Computing System (Originally spelled ‘UNICS’)
Wintel	Windows Operating System on an Intel Machine
Wireshark	A packet sniffer