Payload-based Anomaly Detection in HTTP Traffic

A Thesis submitted for the degree of

Doctor of Philosophy

By

Aruna Jamdagni

In

Faculty of Engineering and information Technology

School of Computing and Communications

UNIVERSITY OF TECHNOLOGY, SYDNEY AUSTRALIA

SUBMITTED NOVEMBER 2012

UNIVERSITY OF TECHNOLOFY, SYDNEY

SCHOOL OF COMPUTER AND COMMUNICATIONS

The undersigned hereby certify that they have read this thesis entitled "**Payload-based Anomaly Detection in HTTP Traffic"** by **Aruna Jamdagni** and that in his opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Doctor of Philosophy.

Principal Supervisor	Co-Supervisor
Prof. Xiangjian (Sean) He	Dr. Priyadarsi Nanda

CERTIFICATE OF AUTHORSHIP/ORIGINALITY

I certify that the work in this thesis has not been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature of Author

Abstract

Payload-based Anomaly Detection in HTTP Traffic

Internet provides quality and convenience to human life but at the same time it provides a platform for network hackers and criminals. Intrusion Detection Systems (IDSs) have been proven to be powerful methods for detecting anomalies in the network. Traditional IDSs based on signatures are unable to detect new (zero days) attacks. Anomaly-based systems are alternative to signature based systems. However, present anomaly detection systems suffer from three major setbacks:

(a) Large number of false alarms,
(b) Very high volume of network traffic due to high data rates (Gbps), and
(c) Inefficiency in operation.

In this thesis, we address above issues and develop efficient intrusion detection frameworks and models which can be used in detecting a wide variety of attacks including web-based attacks. Our proposed methods are designed to have very few false alarms. We also address Intrusion Detection as a Pattern Recognition problem and discuss all aspects that are important in realizing an anomaly-based IDS.

We present three payload-based anomaly detectors, including Geometrical Structure Anomaly Detection (GSAD), Two-Tier Intrusion Detection system using Linear Discriminant Analysis (LDA), and Real-time Payload-based Intrusion Detection System (RePIDS), for intrusion detection. These detectors perform deep-packet analysis and examine payload content using *n*-gram text categorization and Mahalanobis Distance Map (MDM) techniques. An MDM extracts hidden correlations between the features within each payload and among packet payloads. GSAD generates model of normal network payload as geometrical structure using MDMs in a fully automatic and unsupervised manner. We have implemented the GSAD model in HTTP environment for web-based applications.

For efficient operation of IDSs, the detection speed is a key point. Current IDSs examine a large number of data features to detect intrusions and misuse patterns. Hence, for quickly and accurately identifying anomalies of Internet traffic, feature reduction becomes mandatory. We have proposed two models to address this issue, namely two-tier intrusion detection model and RePIDS.

Two-tier intrusion detection model uses Linear Discriminant Analysis approach for feature reduction and optimal feature selection. It uses MDM technique to create a model of normal network payload using an extracted feature set.

RePIDS uses a 3-tier Iterative Feature Selection Engine (IFSEng) to reduce dimensionality of the raw dataset using Principal Component Analysis (PCA) technique. IFSEng extracts the most significant features from the original feature set and uses mathematical and graphical methods for optimal feature subset selection. Like two-tier intrusion detection model, RePIDS then uses MDM technique to generate a model of normal network payload using extracted features.

We test the proposed IDSs on two publicly available datasets of attacks and normal traffic. Experimental results confirm the effectiveness and validation of our proposed solutions in terms of detection rate, false alarm rate and computational complexity.

Acknowledgement

This research would not have been possible without the guidance and the help of many people. First and foremost, my utmost gratitude to my supervisor, Prof. Xiangjian He, for his excellent guidance, support and steadfast encouragement that I will never forget. His comments and suggestions during preparation of this thesis have been extremely valuable. Without his support and supervision, I could not have come this far. I would thank him for his helpfulness and to have been by far more than a simple supervisor.

I would like to express my gratitude to my research co-supervisors, Dr. Priydarsi Nanda and Dr. Ren Liu, for their friendly guidance and unfailing support. Their encouragement has kept me moving ahead at a critical time. Without their help, I would not have been able to complete this thesis.

I appreciate the financial assistance of Australian Postgraduate Award (APA) provided by Australian government and Top-up scholarship provided by the Commonwealth Scientific and Industrial Research Organisation (CSIRO).

Many thanks to my Employer University of Western Sydney and Prof. Simeon Simoff, Dean, School of Computing and Engineering, who gave me time off from work. I will be always grateful for that.

I also appreciate Dr. Qiang Wu and Dr. Wenjing Jia for providing helpful suggestions. My special thanks to collaborator and my good friend Thomas Tan for brain storming discussions. My friends: Thomas Tan and Sheng Wang, they are always helpful whenever I have questions not only on research but also on other matters. It would have been a lonely lab without them.

Last but not the least, I would like to express my love and gratitude to my family members, especially my daughter Divya, my husband Rishi, my sister Meera and brother in-law Satya for their endless love, understanding and encouragement to work on this thesis.

“What we are is God's gift to us. What we become is our gift to God.”

Eleanor Powell

Dedicated to Dear God

Table of Contents

List of Tables

Table of Figures

Acronyms and Abbreviations

ABS	Anomaly Based System
DARPA	Defense Advanced Research Projects Agency
DDoS	Distributed Denial of Service
DoS	Denial of Service
IDES	Intrusion Detection Expert System
IDS	Intrusion Detection System
GATECH	Georgia Institute of Technology
GSAD	Geometrical Structure Anomaly Detection System
GSPM	Geometrical Structure Payload Model
HIDS	Host-based Intrusion Detection System
HTTP	Hyper Text Transport Protocol
IFSEng	Iterative Feature Selection Engine
IDPS	Intrusion Detection and Prevention System
KDD	Knowledge Discovery in Databases
LDA	Linear Discriminant Analysis
LDM	Linear Discriminant Module
McPAD	Multi classifier Payload Based Anomaly Detection
MD	Mahalanobis Distance
MDM	Mahalanobis Distance Map
MIT	Massachusetts Institute of Technology
MS-SQL	MiscroSoft Structured Query Language
NIDS	Network Intrusion Detection System
PA	Parallel Analysis
PAYL	Payload Based Anomaly Detection System
PCA	Principal Component Analysis
PC	Principal Component
RePIDS	Real-time Payload-based Intrusion Detection System
R2L	Remote to Local

SBS	Signature Based System
SRI	Stanford Research International
SVM	Support Vector Machines
TC	Text Categorization
U2R	User to Root

Authors Publications for the Ph.D

Published papers

Journal Papers

1. **A. Jamdagni**, Z. Tan, P. Nanda, X. He, R. Liu, "RePIDS: a Multi Tier Real-Time Payload-Based Intrusion Detection System," *Computer Networks* (ERA **Tier A**), Elsevier, accepted (Final) on 8 October 2012.
2. **A. Jamdagni,** Z. Tan, P. Nanda, X. He, R. Liu, "Mahalanobis Distance Map Approach for Anomaly Detection of Web-Based Attacks," *Journal of Network Forensics*, 2(2), 25-39, 2011.

Conference papers

3. Z. Tan,, **A. Jamdagni**, P. Nanda, X. He, R. Liu, "Network Intrusion Detection based on LDA for payload feature selection," IEEE Globecom 2010 Workshop on Web and Pervasive Security (WPS 2010), Miami, USA, 2010, pp.1590-1594.
4. Z. Tan, **A. Jamdagni**, X. He, P. Nanda, R. Liu, W. Jia, W. Yeh, "A Two-Tier System for Web Attack Detection Using Linear Discriminant Method," Information and Communications Security, LNCS, Vol. 6476/2010, pp.459-471.
5. **A. Jamdagni,** Z. Tan, X. He, P. Nanda, R. Liu, "Mahalanobis Distance Map Approach for Anomaly Detection of Web-Based Attacks," in 8th Australian Information Security Management Conference. November 2010. Perth.

6. **A. Jamdagni,** Z. Tan, P. Nanda, X. He, R. Liu, "Intrusion detection using GSAD model for HTTP traffic on web services," in IWCMC' 10 Proceedings of the 6th International Wireless Communications and Mobile Computing Conference 2010. France: ACM.
7. **A. Jamdagni,** Z. Tan, R. Liu, P. Nanda, X. He, "Pattern Recognition Approach for Anomaly Detection of Web-based Attacks," in the Seventh Annual CSIRO ICT Centre Science and Engineering Conference, November 2010.
8. **A. Jamdagni,** Z. Tan, P. Nanda, X. He, R. Liu, "Intrusion Detection Using Geometrical Structure," in 4th International Conference on Frontier of Computer Science and Technology (FCST 2009), Shanghai, China, December 17-19, 2009, pp. 327-333.
9. **A. Jamdagni,** Z. Tan, R. Liu, P. Nanda, X. He, "A Framework for Geometrical Anomaly Detection Model," in the Sixth Annual CSIRO ICT Centre Science and Engineering Conference, November 2009.