Ain Shams University

Faculty of Engineering

Computer and Systems Engineering Department

Classification of IDS Alerts using Data Mining

A Thesis Submitted in Partial Fulfillment of the Requirements of the Master of Science in Computer and Systems Engineering

Submitted By

Hany Nashat Gabra

B. Sc. In Electrical Engineering

Department of Electrical and Power Engineering

Faculty of Engineering - Cairo University

Under Supervision of

Prof. Dr. Hoda Korashy Mohammed

Department of Computer and Systems Engineering

Faculty of Engineering, Ain Shams University

Ass. Prof. Dr. Ayman Bahaa

Department of Computer and Systems Engineering

Faculty of Engineering, Ain Shams University

Cairo – 2013

Acknowledgements

First, thanks to God who made me able to accomplish this work, I sincerely express my deepest gratitude to my thesis supervisors, Dr. Huda Korashy, Professor of Computer Engineering, Faculty of Engineering, Ain Shams University, and Dr. Ayman Bahaa, Associate Professor of Computer Engineering, Faculty of Engineering, Ain Shams University.

I like to express my thankfulness for their kind supervision and offering unfailing support, invaluable advices and comments and helpful and useful discussions in selecting the interesting point and during the preparation of this thesis. I owe a special acknowledgment to them for giving me a lot of their time during the years of preparing this thesis. I could never had done it without their support, technical advice and suggestions, thorough reading of all my work.

I would like to thank Dr.Ayman Taha who provide me with data during his data collection session at the Center for Education and Research of Information Assurance and Security (CERIAS), Purdue University, USA.

Finally, I am very grateful to my wife Nevine, and my son for their sacrifices, sustained moral support, and encouragement. I always thank my God for blessing me with such a wonderful family.

Abstract

Intrusion detection systems (IDSs) have become a widely used measure for security, but we still have a problem on these systems results which include many irrelevant alerts.

IDS is a security measure for network monitoring and protection. Unfortunately, IDSs are known to generate large amounts of alerts, with many of them being either false positives or of low importance. It is too hard for the human to spot alerts which need more attention. In order to tackle this issue, we have proposed an IDS alert classification method based on data mining techniques in order to distinguish serious alerts and irrelevant.

A data mining method has been used in our classification method to distinguish serious alerts and irrelevant one. The performance of 99.9 % has been reached in comparison with the other recent data mining methods that shown that they have reached the performance of 97% maximum.

Also, a new ranking technique has been proposed to order the alerts in a list according to the alert's importance to minimize the human interventions and to give the opportunity to investigate the important alert before the less important one that will be pushed automatically to the end of the proposed alert list.

اصبح استخدام نظم الانذارات وكشف الاختراقات مستخدما بشكل واسع في التأمينية في الاونة الاخيرة ولكن هناك مشاكل تتمثل في وجود نسبة الانظمة كبيرة من الانذارات الزائفة التي تؤثر في نتائج تلك النظم.

تعتبر نظم الانذارات وكشف الاختراقات من وسائل تأمين الشبكات ومراقبتها ولكن للاسف تتنتج هذه الانظمة كم هائل من الانذارات الزائفة او القليلة الاهمية مما يصعب المهمة علي مراقبين الشبكات بمراجعة هذا الكم الهائل من الانذارات واكتشاف الانذارات التي تستدعي التدقيق.

للتغلب علي هذه المشكلة تم تقديم اطروحتنا التي تعتمد علي تطوير وسائل لتصنيف هذه الانذارات اعتمادا علي تطوير اليات التنقيب في البيانات.

ويظهر في اطروحتنا اننا تمكنا من فصل الانذارات المزيفة بنسبة تزيد عن 99% وهذا يزيد عن النتائج التي تم الوصول اليها في ابحاث سابقة والتي بالاضافة الي ما سبق فقد تم فرض طرق وصلت الي 97% بحد اقصي.

الترتيب قائمة الانذارات بناءا علي اهمية كل انذار وبذلك نعطي الفرصة لمراقي الشبكات بمراجعة الانذارات بناءا علي اهميتها ونمكنهم من التركيز علي حل الانذارات الفعلية وعدم اضاعة وقته في الانذارات الاقل اهمية والتي تأتي في اخر القائمة بشكل تلقائي.

Keywords:

Intrusion Detection, Frequent Pattern, Frequent Itemset, support, minisupport

Statement

This dissertation is submitted to Ain Shams University for the degree of Master of Philosophy in Computer and Systems Engineering Department.

The work included in this thesis was carried out by the author at Computer and Systems Engineering Department, Faculty of Engineering, Ain Shams University.

No part of this thesis has been submitted for a degree or qualification at other university or institution.

Date : / / 201
Signature :
Name : Hany Nashat Gabra

Table of Contents

List of Figures:

LIST OF TABLES:

List of Equations:

LIST OF ABBREVIATION:

IDS : Intrusion Detection system

Minisupport : minimum support

MinConf: minimum confidence

Performance: IDS performance or the main alerts reduction rate.

FPOF: Frequent Pattern Outlier Factor

FPS: Frequent Patterns

Classification of IDS Alerts using Data Mining Techniques

Chapter One

Introduction

1.1 General

Intrusion detection systems IDS is a security measure for network monitoring and protection. Unfortunately, IDSs are known to generate large amounts of alerts, with many of them being either false positives or low importance.

This makes it hard for the human to spot alerts which need more attention. In order to tackle this issue; we have proposed an IDS alert classification method based on data mining techniques

An important problem is still not fully addressed thus IDS can produce a large number of alerts which are overwhelming to the human. For example, a single IDS sensor can generate tens of thousands of alerts in a day [16, 20].

Furthermore, vast majority of the alerts are false positives or of low importance [8,16].

Intrusion detection systems (IDSs) have become a widely used measure for security. It is not unusual to receive thousands of alerts from a single network IDS sensor per day, with more than 90% of the alerts being irrelevant [1, 2, 3].

In order to distinguish important IDS alerts from irrelevant one, IDS alert log analysis techniques are often used. Many techniques have been suggested for this purpose like machine learning [4], time series modeling [3, 5], the use of control charts [6], etc. During the last decade, data mining based technique have also been proposed in a number of research papers [1, 2, 7, 8, 9, 10]. With these techniques, IDS alert logs from the recent past are mined for knowledge that is used for the creation of event filtering and correlation rules for future IDS alerts. However, existing methods are inherently semi-automated – they assume that a human expert interprets detected knowledge and creates event filtering and correlation rules by hand.

We have propose a data mining based method for classification to distinguish serious alerts and irrelevant one and re-list the alerts according to its importance.

1.2 IDS Terminology

An alert or an alarm is defined as a signal reporting that a system has been, or is being, attacked. A True Positive or True attack alert is defined as the case when a real attack triggers IDS to produce an alarm; this alarm is a true attack or correct alarm, the IDS reduction rate of the original alerts produced by the IDS is the main IDS performance presented by the following equation:

IDS performance = True attacks / (True attacks + Low importance)...(1.1)

For example, IDS that produces 100 alerts for 10 real attacks and other 90 non-attack alerts. This situation can be expressed as: IDS Alerts: 100, True attacks: 10 alerts, Irrelevant: 90 alerts

By using equation 1.1 the IDS performance = 10 / (10+90) = 10 %

1.3 Methodology

This research has been done through several steps:

1. Existed IDS systems for the IDS and different data mining techniques. This study allowed us to identify the missing part and the drawback of those systems.
2. Introduce modification to overcome the problem by presenting the idea that will be used to solve the current IDS system problems and improve its accuracy.
3. Investigate a datasets and perform an experimental session to test the proposed idea.

Our dataset used on our implementations was collected on June 22 2010 by Dr.Ayman Taha during his data collection session at the Center for Education and Research of Information assurance and security (CERIAS), Purdue University, USA.

The collected data included real and attempted attacks that have been used to implement a method to prove the theory of the idea and to assure the success of the model.

Finally, we extract the method results and perform the needed analysis for these results for the purpose of assessment and presentation

1.4 Contributions

This dissertation provides solutions for the problems outlined above, and provides the following contributions:

- Enhance accuracy and completeness of intrusion detection;
- Improve the ability to distinguish serious attacks from less important ones;
- Safe the security analyst / human effort by sorting the alerts according to their importance.

1.5 Dissertation Organization

The remainder of this dissertation is structured as follows, Chapter 2 presents a survey of intrusion detection and related work, describing and giving an introduction to the current IDS.

Chapter 3,An introduction survey to current data mining techniques and algorithms.

Chapter 4 introduces the work has been done for our enhanced mining tool for intrusion detection.

In Chapter 5, detailed experimental results of applying the proposed model on the dataset are presented.

The conclusion and future work are introduced in chapter 6.

Chapter Two
Intrusion detection Survey

In this chapter, a literature survey of intrusion detection techniques is presented.

2.1 The Importance of Security and Intrusion Detection

In the past the computer systems usually were not networked, or connected to a small network spanning a company or a building. Today almost every computer system is connected to the Internet.

The main concern with this situation is that the number of potential attackers that can attack a given system has been increased drastically.

Whereas before an attacker had to be physically present at the console of the computer or be connected to the same local area network as the target computer, today's attacker can be located almost anywhere in the world [48].

Another reason for the increased importance of computer security is that more sensitive data are stored on computers than before. For instance medical records and bank accounts are not paper based anymore.

Another change that has been happened lately is that many businesses rely on computer systems to perform their function. As a result, a computer system problem can shut down the whole system.

For instance, a web based store would not get any customers if their network connection failed.

All these changes in data processing now a day in business increases the number of potential targets for attacks and the effect of successful attacks have become more serious. In addition, the attackers have improved their attack techniques.
It is now common to see large scale coordinated attacks where the attacker utilizes efforts in order to attack a single target.
These kinds of attacks can be challenging to defend against, as it is not easy to identify the attacker when he is using multiple hosts. In addition, the attacker's computers are often located in different networks [49].

This potentially makes the aggregate network throughput to be available to the attacker very large and can enable the attacker to flood the victim's network with traffic, creating a denial-of-service attack.

Therefore, computer security is an increasingly important topic. It is important to insure that the secrecy of sensitive data is protected, the integrity of important data is not violated, and the availability of critical systems is guaranteed. Computer security tries to achieve all these goals [49].