Egyptian Informatics Journal (2014) **15**, 13–24

Cairo University

Egyptian Informatics Journal

www.elsevier.com/locate/eij
www.sciencedirect.com

ORIGINAL ARTICLE

A proposed HTTP service based IDS

Mohamed M. Abd-Eldayem *

IT Department, Faculty of Computers and Information, Cairo University, Egypt
CEN Department, College of Computers and Information Sciences, King Saud University, Saudi Arabia

Received 9 October 2013; revised 1 January 2014; accepted 6 January 2014
Available online 30 January 2014

KEYWORDS

Computer security;
Network security;
Intrusion Detection System (IDS);
Naïve Bayes classifier

Abstract The tremendous growth of the web-based applications has increased information security vulnerabilities over the Internet. Security administrators use Intrusion-Detection System (IDS) to monitor network traffic and host activities to detect attacks against hosts and network resources. In this paper IDS based on Naïve Bayes classifier is analyzed. The main objective is to enhance IDS performance through preparing the training data set allowing to detect malicious connections that exploit the http service. Results of application are demonstrated and discussed. In the training phase of the proposed IDS, at first a feature selection technique based on Naïve Bayes classifier is used, this technique identifies the most important HTTP traffic features that can be used to detect HTTP attacks. In the testing and running phases proposed IDS classifies the network traffic based on the requested service, then based on the selected features Naïve Bayes classifier is used to analyze the HTTP service based traffic and identifies the HTTP normal connections and attacks. The performance of the IDS is measured through experiments using NSL-KDD data set. The results show that the detection rate of the IDS is about 99%, the false-positive rate is about 1%, and the false-negative rate is about 0.25%; therefore, proposed IDS holds the highest detection rate and the lowest false alarm compared with other leading IDS. In addition, the proposed IDS based on Naïve Bayes is used to classify network connections as a normal or attack. And it holds a high detection rate and a low false alarm.

1. Introduction

These days, most of the universities, organizations and companies provide their services through a computer network and Internet technologies; therefore, most of their tasks are accomplished using web-based applications. However, the attackers could have exploited the Internet to break into the local network to gain the confidential information or to compromise the network resources. Some security tools such as firewalls, anti-virus software and Intrusion-Detection Systems (IDSs) are used to protect the network and to thwart hackers. IDSs are used to monitor network traffic to detect the intruder

* Address: CEN Department, College of Computers and Information Sciences, King Saud University, Saudi Arabia. Tel.: +966 592626083; fax: +966 014676990.
E-mail address: mdayem@gmail.com

Peer review under responsibility of Faculty of Computers and Information, Cairo University.

ELSEVIER Production and hosting by Elsevier

1110-8665 © 2014 Production and hosting by Elsevier B.V. on behalf of Faculty of Computers and Information, Cairo University.
http://dx.doi.org/10.1016/j.eij.2014.01.001

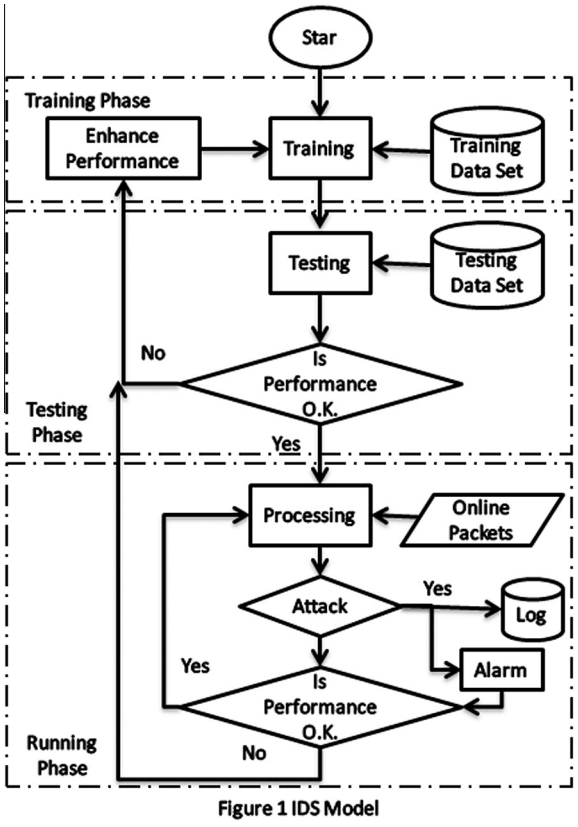

Figure 1 IDS model.

network events. Statistical methods are utilized widely in the IDSs since these methods are analytical methods that depend on the probabilistic, and they are considered as the major tools that are coping with uncertain data. Many of these methods are used to enhance the IDS performance; they give a better way to classify network events as normal or as attacks. Bayesian classifier is an example of the statistical methods that are used in IDSs. In the training phase, it computes the conditional probabilities for each of the normal and the attack classes. It uses a training data set that is classified into attack and normal classes. And in the testing and running phase the classifier uses these probabilities to extract the belongings probabilities of different classes; therefore, the unknown network traffic can be classified as the class that supposes maximum value [1,2]. This means that the running network traffic would be classified as normal or as attack events. In this paper, IDS based on Naïve Bayes classifier is proposed and its performance is tested through practical experiments. The following sections of this paper are organized as follows: the second section summarizes the ideas of the leading IDSs. Section 3 describes the proposed IDSs. The implementations, discussions and experimental results for these IDSs are described and illustrated in Section 4, finally the conclusions and future works are summarized in Section 5.

2. Related works

In [3] anomaly detection IDS is proposed, it combines k-Means, K-nearest neighbor classifier and Naive Bayes classifier. It selects the important features using an entropy based algorithm, then it applies k-Means in clusters' phase. It is not only able to detect the attacks, but also it can classify them into four types: DOS, U2R, R2L and probe. This system can achieve 98.18% detection rate and 0.83% false-positive rate; however, it may increase the processing time because it executes the k-Means, K-nearest neighbor classifier and Naive Bayes classifier. In addition, the accuracy of classifying the attack types ranges from 92% to 98%. A score-based multi-cycle detection algorithm based on Shiryaev–Roberts procedure is proposed in [4]. Comparing to similar procedures; Shiryaev–Roberts procedure is computationally inexpensive in addition it is easy to be practically applied in real time IDS. The target

Table 1 List of features of a record in KDD data set.

Index	Feature	Index	Feature
1	duration	22	is_guest_login
2	protocol_Type	23	count
3	service	24	srv_count
4	flag	25	serror_rate
5	src-bytes	26	srv_serror_rate
6	dst_bytes	27	rerror_rate
7	land	28	srv_error_rate
8	wrong_fragment	29	same_srv_rate
9	urgent	30	diff_srv_rate
10	hot	31	srv_diff_host_rate
11	num_failed_logins	32	dst_host_count
12	logged_in	33	dst_host_srv_count
13	num_compromised	34	dst_host_same_srv_rate
14	root_shell	35	dst_host_diff_srv_rate
15	su_attempted	36	dst_host_same_src_port_rate
16	num_root	37	dst_host_srv_diff_host_rate
17	num_file_creations	38	dst_host_serror_rate
18	num_shells	39	dst_host_srv_serror_rate
19	num_access_files	40	dst_host_rerror_rate
20	num_outbound_cmds	41	dst_host_srv_error_rate
21	is_hot_login		