ISSN 1746-7659, England, UK
Journal of Information and Computing Science
Vol. 9, No. 1, 2014, *pp*. 037-048

A Novel Data Mining based Hybrid Intrusion Detection Framework

Mradul Dhakar[+] and Akhilesh Tiwari

Department of CSE & IT, Madhav Institute of Technology and Science,
Gwalior (M.P.), India

(*Received June 23, 2012, accepted October 12, 2013*)

Abstract. The prosperity of technology worldwide has made the concerns of security tend to increase rapidly. The enormous usage of internetworking has raised the need of protecting system(s) as well as network(s) from the unauthorized access (intrusion). To tackle the intrusive activities, several countermeasures have been found in literature viz. firewall, antivirus and currently widely preferred Intrusion detection System (IDS). IDS, is a detection mechanism for detecting the intrusive activities hidden among the normal activities. The revolutionary establishment of IDS has attracted analysts to work dedicatedly enabling the system to deal with technological advancements. Hence in this regard, various beneficial schemes and models have been proposed in order to achieve enhanced IDS. This paper proposes a novel hybrid model for intrusion detection. The proposed framework in this paper may be expected as another step towards advancement of IDS. The framework utilizes the crucial data mining classification algorithms beneficial for intrusion detection. The Hybrid framework would henceforth, will lead to effective, adaptive and intelligent intrusion detection.

Keywords: Data Mining, Intrusion Detection, Classification, K2, TAN, REP, KDDCup'99

1. Introduction

The progressive use of intrusion detection system for handling the abnormalities on web has caused multiple efforts laid by the analysts. The intrusions have been found dominating the internet which may be assumed as a threat to the security of genuine users. In order to meet the advancement of changing technological world, IDS has been through various alterations where it has been competent in detecting these intrusions more precisely.

Though IDS itself is a standalone definition but in order to make the system cope with the recent technological developments and increased intrusions strategies, it has gone through several revisions where it has beneficially used the various research fields such neural network, statistics and recently data mining. The key ideas are to use data mining techniques to discover consistent and useful patterns of system features that describe program and user behavior, and use the set of relevant system features to compute (inductively learned) classifies that can recognize anomalies and known intrusions [1]. This thought has made IDS with data mining serve as the most promising intrusion detection scheme where Data Mining identifies trends within data that go beyond simple analysis [2].

Generally IDSs are deployed to monitor a system or a network in search of any abnormal condition. In this surveillance if any kind of intrusive attempt is detected, the monitoring system i.e. IDS sets up an alarm which is an indication of the presence of intrusion. In order to detect intrusions in an efficient manner, various appreciable models have registered their presence in the literature. The presently available models involve usage of various novel algorithms which are likely to detect these intrusions distinguishably.

Among these, algorithms based on data mining have been a point of attraction for researchers because of their extensive feasibility in detecting intrusions. These algorithms aid in improving accuracy of the system along with effective detection rate and less false alarm rate. The algorithms loyal for classification are the

[+] Corresponding author.
E-mail address: mraduliitm@gmail.com.

most desirable algorithms for detection.

In the data mining classification techniques, Tree Augmented Naïve Bayes (TAN) and Reduced Error Pruning (REP) algorithms have come out as the most significant detection algorithms in IDS. Hence this paper presents an intelligent effort for intrusion detection which proposes a framework named Hybrid Intrusion Detection Model. This model is a combinational scheme which aims at surmounting the shortcomings faced by two algorithms individually with interestingly increased accuracy of the detection.

The paper consists of following sections: Section II is a brief description about intrusion detection system, section III discusses the intrusion detection processes involved, section IV studies about the intrusion detection approaches, section V describes about the attacks detected by IDS, section VI elaborates the proposed methodology and section VII consists the experimental analysis performed on proposed hybrid model.

2. Intrusion Detection System

The adverseness of abnormalities (generally referred as intrusions) on web has brought up the security concerns leading to the successful implementation of abnormalities detection system named as Intrusion Detection System (IDS). Intrusions may be defined as the unauthorized attempt for gaining access on a secured system or network. Intrusion detection is the course of action to detect suspicious activity on the network or a device. Intrusion Detection System (IDS) is an important detection used as a countermeasure to preserve data integrity and system availability from attacks [3].

The IDS has been a renowned aspect for detecting intrusions adequately. The IDS is assumed as hardware or software or combination of both that allows monitoring of the network traffic in search of intrusions. An intrusion detection system (IDS) inspects all inbound and outbound network activity and identifies suspicious patterns that may indicate a network or system attack from someone attempting to break into or compromise a system [4]. It has advantageously helped the analysts to learn about the various possible attacks.

2.1. Intrusion Detection Process

Intrusion detection on the basis of their detection process are categorize into Misuse / Signature-based intrusion detection and Anomaly-based intrusion detection.

2.1.1. Misuse Detection

Misuse detection compares the user activities to the known intruder activities on web. The idea of misuse detection is to represent attacks in the form of a pattern or a signature so that the same attack can be detected and prevented in future [5]. The IDS searches for defined signatures and if a match is found, the system generates an alarm indicting the presence of intrusion. Since it works on the basis of predefined signatures, it is unable to detect new or previously unknown intrusions.

2.1.2. Anomaly Detection

Anomaly intrusion detection identifies deviations from the normal usage behavior patterns to identify the intrusion [6]. It is a technique which is based on the revealing of traffic anomalies. It estimates the deviation of a user activity from the normal behavior and if the deviation goes beyond a preset threshold, it considers that activity as an intrusion. It is because of this threshold concept anomaly can detect new intrusions in addition to the previously known intrusions. However anomaly is able to detect new intrusion but the compulsion for involvement of limiting factor results in high percentage of false positive rate.

2.2. Intrusion Detection Approaches

On the basis of the data analyzed and stored it is classified into Host-based Intrusion Detection System (HIDS) and Network-based Intrusion Detection System (NIDS).

2.2.1. Host-based Intrusion Detection System