Available online at www.sciencedirect.com

ScienceDirect

Journal of Electrical Systems and Information Technology xxx (2016) xxx–xxx

www.elsevier.com/locate/jesit

Machine learning approach to detect intruders in database based on hexplet data structure

Q1 Saad M. Darwish

Department of Information Technology, Institute of Graduate Studies and Research, Alexandria University, 163 Horreya Avenue, El-Shatby 21526, P.O. Box 832, Alexandria, Egypt

Received 15 July 2015; received in revised form 30 October 2015; accepted 15 December 2015

Abstract

Most of valuable information resources for any organization are stored in the database; it is a serious subject to protect this information against intruders. However, conventional security mechanisms are not designed to detect anomalous actions of database users. An intrusion detection system (IDS), delivers an extra layer of security that cannot be guaranteed by built-in security tools, is the ideal solution to defend databases from intruders. This paper suggests an anomaly detection approach that summarizes the raw transactional SQL queries into a compact data structure called hexplet, which can model normal database access behavior (abstract the user's profile) and recognize impostors specifically tailored for role-based access control (RBAC) database system. This hexplet lets us to preserve the correlation among SQL statements in the same transaction by exploiting the information in the transaction-log entry with the aim to improve detection accuracy specially those inside the organization and behave strange behavior. The model utilizes naive Bayes classifier (NBC) as the simplest supervised learning technique for creating profiles and evaluating the legitimacy of a transaction. Experimental results show the performance of the proposed model in the term of detection rate.

Keywords: Database security; Anomaly detection; Database intrusion detection; Role-based profiling

1. Introduction

In today's business realm, the most valuable asset of organizations is its information and thus needs efficient management and protection. Over the last few years, database systems constitute the central of the information systems infrastructure because they permit the efficient administration and retrieval of huge amounts of data in addition to offer mechanisms that can be employed to certify the integrity of the stored data. Data found in these databases vary between private information, banking transactions, personal medical data, and commercial contracts, etc. Any violation

E-mail address: saad.darwish@alex-igsr.edu.eg

Peer review under the responsibility of Electronics Research Institute (ERI).

ELSEVIER Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.jesit.2015.12.001

of security to these databases will lead to passing reputation of the organization, loss of customers' sureness and might even lead to lawsuits.

There are many ways to secure databases like user authentication, transaction data encryption, data watermarking, and intrusion detection. Each of these ways has its benefits and they should work together to reach the maximum security level of databases. User authentication is a prevention technique that prevents unauthorized users from gaining access to the database. Transaction data encryption is a prevention technique also; which thwarts attackers from understanding the data in case of sniffing on the session. Watermarking the data is a detection technique that used to pledge data integrity. Intrusion Detection is a detection technique that is used to identify the malicious activities as early as possible if the system prevention mechanisms were bypassed to minimize the harm caused by intruders; for more information regarding these techniques readers can refer to Panda and Giordano (1999) and Agrawal and Kiernan (2002).

Actually, available database security mechanisms are not basically designed to detect intrusions; they are intended to avoid the intruders. So, intrusion detection system is considered to be the second defense line. Database Intrusion is commonly defined as a set of actions that try to violate data integrity, data confidentiality or data availability; while database intrusion detection is the process of tracking transactions submitted to database and analyzing them to detect the possible presence of intruders. In general, there are two types of database intrusion attacks (I) insider and (II) outsider (Panda and Giordano, 1999). The insider's attacks are the ones when an intruder has all the privileges to access the database but he performs malicious actions. The outsider's attacks are the ones when the intruder does not have proper rights to access the database. He attempts to first rush into and then performs malicious actions. Detecting insider attacks are often more difficult than detecting outsider's attacks.

In the literature, database intrusion detection has two main models (Agnew, 2003; Chandola et al., 2009), anomaly detection and misuse detection. The anomaly detection model is based on the profile of a user's normal behavior. It analyzes a user's current session and compares it against the profile representing his normal behavior. If a major deviation is found during the comparison, an alarm will be raised. This includes monitoring the system states over a period of time; with assume that this monitored profile can mark the "normal" profile of the system. In general, it is not easy to implement an anomaly detection solution because profiling "normal" usage is a challenging task and anomaly detection methods are also subjected to high false-positive rates, particularly with over-fitted "normal" profiles. On the other hand, relaxing the margins on these profiles might result in missed attacks. In this case, fine-tuning anomaly detection systems to find the optimum thresholds is a major issue. In contrast, the misuse detection model is based on comparing user's session or commands with the signature of attacks previously used by attackers. So, the signature of attacks should be well known to be detected in this model inside regarding data. Misuse detection is clever to detect only known attacks. However, misuse detection can sometimes detect new attacks which share characteristics with previously known attacks.

The possibility of enhancing existing anomaly detection systems by presenting an effective database IDS constitutes the objective of this work. The contribution of this paper is an anomaly detection system explicitly adapted to the RBAC database system. This system builds and maintains a role profile representing precise and consistent user behavior by means of special data structure that is able to determine role intruders. The recommended data structure is transaction-based which extracts relationship among queries in the same transaction instead of the state-of-the-art approaches that is query-based approach that can detect the attributes which are to be referred together but cannot detect the queries which are to be executed together.

The paper is organized as follows: next section presents some of state-of-the art work in detecting anomalous database requests. Then Section 3 discusses in detail the proposed work in this field. Section 4 reports the related experimental results and then Section 5 end with concluding the paper and demonstrate some preliminary ideas for future work.

2. Related work

In recent years, researchers have proposed a variety of approaches for increasing the intrusion detection efficiency and accuracy. But most of these efforts concentrated on detecting intrusion at the network or operating system level (Kang et al., 2005; Ghosh et al., 1999). For instance, the technique presented in Kang et al. (2005) attempts to profile normal behavior for a program. It tries to conclude the normal system call sequences and save these sequences in the database as normal program access patterns. These efforts, however, are not adequate for protecting the database and