Chapter 62
A Proposal for Information Systems Security Monitoring Based on Large Datasets

Hai Van Pham
Hanoi University of Science and Technology, Hanoi, Vietnam

Philip Moore
Lanzhou University, Lanzhou, P.R. China

ABSTRACT

This article describes how the objective of recent advances in soft computing and machine learning models is the resolution of issues related to security monitoring for information systems. Most current techniques and models face significant limitations, in the monitoring of information systems. To address these limitations, the authors propose a new model designed to detect potential security breaches at an early stage using logging data. The proposed model uses unsupervised training techniques with a rule-based system to analyse data file logs. The proposed approach has been evaluated using a case study based on the learning of data file logs to determine the effectiveness of the proposed approach. Experimental results show that the proposed approach performs well, the results demonstrate that the proposed approach performs better than other conventional security methods in the identification of the correct decisions related to potential security in information systems.

INTRODUCTION

In considering Information Systems (IS) security, the detection of potential security breaches plays an important role in protecting organizations from both external and internal attacks. Viewed from a security perspective, logging data forms an important component in the range of tools available in IS security analysis. Analysis of data logs has the potential to greatly increase an organization's understanding of an insider's behaviour or malicious activity occurring across the network.

DOI: 10.4018/978-1-7998-7705-9.ch062

In data file logs, event correlation services and server processes provide many unstructured data sets related to events (both normal and security related) and security alarms for use by security analysts. In the analysis of data logs, there are two types of analysis tools which are: offline and online monitoring (SourceForge, 2017). In related research (Wurzenberger, 2016; Anastopoulos, 2017) these approaches are applied using clustering and Markov chains to create a synthetic log data. The proposed model requires only a small set of real network data as an input to understand complex 'real system' behaviour. The importance of extending data log files lies in monitoring techniques which implement more complex heuristic approaches and event correlation. In addressing log management, an infrastructure for 'real-time' security monitoring on a large-scale infrastructure is proposed (Mao, 2017) to monitor such systems for security analysis and reporting purposes Further investigation is used to apply statistical modelling designed to capture each host's network behaviour patterns and reduce the complexity of analysis (Suriadi, 2017a, 2017b).

In related research, see: (Ambre, 2015; Abbott, 2015; Wurzenberger, 2016), many studies have proposed approaches for data file log analysis based on historical IS data. Data log analysis and event correlation are closely related to each other in the collection of information for use in the detection of insider threats, for example Dario Forte (Cfe, 2009) has investigated log files in security incident prevention (Casey, 2007; Bonchi, 2001). Juvonen anomaly detection has been proposed (Juvonen, 2015) in an approach based on dimensionality reduction techniques for HTTP log analysis (Zhong, 2011; Bonchi, 2001). Recent research by Xiu-yu Zhong (Zhong, 2011; Shebaro, 2010) has presented an approach designed to realise the monitoring of IS in which data file logs are analysed using data mining techniques; however, there are limitations in this approach in that it uses partial monitoring of offline data sets.

Big data and object analysis provide valuable insights into how patterns of data are related, the aim being to enable 'better-informed' decision-making in IS security (Suriadi, 2017a, 2017b; Dang-Pham, 2017). The investigation of Process-oriented data mining uses algorithms, patterns, and event logs to construct models addressing IS security and user behavioural issues (Dang-Pham, 2017; Bugliesi, 2017). These studies have investigated exponential random graph modelling to predict the occurrence of information security breaches while a range of alternative approaches have been applied to classify and review proposals around formal methods for web security.

In considering the design of IS's, intrusion detection is becoming increasingly important given the need to process and monitor events in 'real-time' in both computerized and networked systems. Analysis of such events entails monitoring to identify potential incidents such as security breaches in the form of violations or imminent threats of violation of computer security policies, acceptable use policies, or standard security practices.

In this paper, we present a design approach for IS security predicated on the identification of potential security breaches at an early stage using the monitoring of data logs has been proposed. Our proposed model is based on the use of Self-Organizing Maps (SOM) using unsupervised training techniques combined with a rule-base. Our objective is to support the monitoring of security (using log files) and decision making in the identification of risk(s) as such risks relate to malicious activity. Our proposed model analyses behaviour or malicious activity for both internal and external attacks. Malicious behaviour or activity, when taken together with unsupervised training in online or offline monitoring (of data log files), aims to analyse groups of malicious activities in IS security monitoring. In considering the proposed security model, the contribution proposed in this study includes addressing the issues identified along with: (a) the capability to enhance the monitoring of IS to improve IS security using mali-