Journal of Network and Computer Applications 72 (2016) 14–27

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

journal homepage: www.elsevier.com/locate/jnca

Review

User profiling in intrusion detection: A review

Jian Peng[a], Kim-Kwang Raymond Choo[b,a,c,*], Helen Ashman[a]

[a] School of Information Technology and Mathematical Sciences, University of South Australia, Australia
[b] Department of Information Systems and Cyber Security, University of Texas at San Antonio, USA
[c] School of Computer Science, China University of Geosciences, Wuhan, China

ARTICLE INFO

Article history:
Received 23 February 2016
Received in revised form
20 June 2016
Accepted 29 June 2016
Available online 1 July 2016

Keywords:
Behavioural biometrics
Intrusion detection and prevention systems
Psychometrics
User behaviour
User profiling

ABSTRACT

Intrusion detection systems are important for detecting and reacting to the presence of unauthorised users of a network or system. They observe the actions of the system and its users and make decisions about the legitimacy of the activity and users. Much work on intrusion detection has focused on analysing the actions triggered by users, determining that atypical or disallowed actions may represent unauthorised use. It is also feasible to observe the users' own behaviour to see if they are acting in their'usual' way, reporting on any sufficiently-aberrant behaviour. Doing this requires a user profile, a feature found more often in marketing and education, but increasingly in security contexts. In this paper, we survey literature on intrusion detection and prevention systems from the viewpoint of exploiting the behaviour of the user in the context of their user profile to confirm or deny the legitimacy of their presence on the system (i.e. review of intrusion detection and prevention systems aimed at user profiling). User behaviour can be measured with both behavioural biometrics, such as keystroke speeds or mouse use, but also psychometrics which measure higher-order cognitive functions such as language and preferences.

Contents

1. Introduction

An intrusion detection system (IDS) monitors host systems and/or network traffic for suspicious activity. Once it finds any, it alerts the system or network administrator. In some cases, the IDS may also respond to anomalous or malicious traffic by taking

* Corresponding author at: Department of Information Systems and Cyber Security, University of Texas at San Antonio, USA.
E-mail address: raymond.choo@fulbrightmail.org (K.-K. Choo).

http://dx.doi.org/10.1016/j.jnca.2016.06.012
1084-8045/© 2016 Elsevier Ltd. All rights reserved.

action such as blocking the user or source IP address from accessing the network.

Intrusion detection systems are generally classified according to where they perform their observations. An IDS can be *network-based* or *host-based*. A network-based IDS observes strategic points within the network to monitor traffic to and from all devices on the network. In contrast, a host-based IDS runs on an individual host or device on the network, monitors the inbound and outbound packets from that device only and alerts the user or administrator if suspicious activity is detected.

Besides these two types of IDS, another proposed by Pennington et al. (2010) is storage-*based* intrusion detection, which analyses all requests received by the storage server and determines the system intrusions by the profiles of data access patterns of systems. As there are lots of logs/traces on storage devices, they can be used for intrusion analysis (Khan et al., 2016). The advantages are that it can be independent from the client's operating systems and continues to identify the intrusions after systems have been compromised, whereas host-based and network-based IDS can comparably be easier be disabled by the intruder; since storage devices are often on different platforms, having restricted interface to outside, it can be more difficult for intruders to compromise them and delete their attack logs and traces which have also been used in forensic investigations This type of IDS are generally used for intrusion detection in storage area network, object-based storage devices, workstation disk drives (Rahman and Choo, 2015; Martini and Choo 2014; Quick et al., 2013; Yampolskiy and Govindaraju, 2008).

IDSs are also often classified according to their primary technique, and can be either *signature-based* (also known as rule-based), or *anomaly-based*. The signature-based IDS monitors packets on the network and compares them against a database of signatures or attributes from known, previously-established malicious threats. This is similar to the way most antivirus software detects malware (Rhodes et al., 2000; Alexandre, 1997; Cortes and Pregibon, 2001; Han et al., 2002; Venugopala and Hu, 2008; Blasing et al., 2010). Although this technique is considered the *de facto* standard, a key limitation is the delay associated with updating the IDS signatures of new intrusions(Afroz et al., 2012), and during that time the IDS is unable to detect the new threat (e.g. zero-day vulnerabilities).

In contrast, the anomaly-based IDS technique is able to detect new forms of attack without prior notification of them. Instead it monitors network traffic and compares it against an established baseline, where the baseline identifies what is “normal” for that network, what protocols are generally used, what ports and devices generally connect to each other. It alerts the administrator or user when anomalous or significantly different traffic is detected (Keselj et al., 2003; Barron-Cedeno et al., 2010; Marceau, 2000; Shrestha and Solorio, 2013; Houvardas and Stamatatos, 2006). However, it may miss both known and novel attacks if they are not manifested along the observed dimensions. Also, depending on how finely-tuned the analysis is, it can have a high error rate, either alerting genuine behaviour as an intrusion (i.e. a false positive) or conversely, not detecting an intruder (i.e. a false negative). Additionally, it needs purity of training data, i.e. an absence of attacks when creating the initial baseline against which to compare later activity. Finally, it is a *post facto* technique which can only detect an attack once it has already occurred, and which may be easy to evade once the model is known.

A typical IDS (Denning, 1987; Mitchell and Chen, 2014; Yeung and Ding, 2002) includes the following components:

1) Data collector collects relevant data from the sensors on monitored devices or systems.
2) Profile generator analyse the data from the Data collector and generate profiles. The anomaly based IDS builds the normal profiles automatically but a signature based IDS may involve experts' efforts to generate its malicious signatures during the training stage.
3) Profile templates store the profiles and are shaded in Fig. 1. A signature based IDS saves malicious profiles (signatures) while an anomaly based IDS saves normal profiles.
4) Intrusion detector is the key component in the IDS and carries out the task of detecting intrusion based on the current profiles.
5) Post processor is responsible proper actions taken once intrusion take place.
6) Profile updater makes proper updates based on current received profiles and relevant algorithms (Fig. 1).

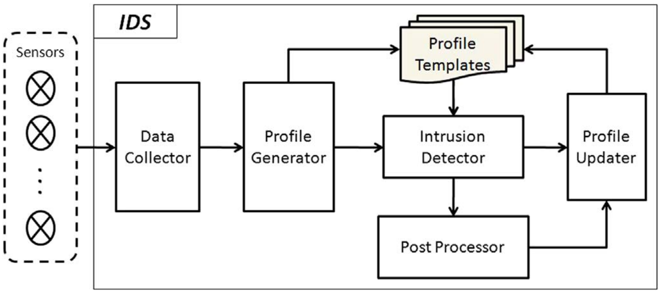

Fig. 1. A typical IDS architecture.

Clearly, there is benefit in operating complementary approaches to intrusion detection, not just in signature-based and anomaly detection techniques, but in a combination of anomaly detection profiles. Keystroke analysis is highly effective for intrusion detection but will throw up false positives if, for example, the user has an arm injury which causes them to type differently. However, combining it with other profiles, such as habitual web sites, favoured applications, normal access time, and so on, will help ameliorate detection errors generated by singular deviations in one profile.

Behavioural science is concerned with gaining a better understanding of human behaviour which focuses specifically on criminal human behaviour in an attempt to better understand criminals—who they are, how they think, why they do what they do—as a means to help solve malicious intrusions. A definition of “behavioural profiling” as offender profiling suggests techniques used to identify likely suspects and analyse patterns that may predict future offences and/or victims (Woodhams, Toye). These techniques are able to help investigators to accurately predict and profile the profiles of unknown criminal subjects or offenders. Behavioural profiling can be either used to identify a potential intruder or to determine normal user patterns, but it is much harder to profile an anomaly behaviour as intruders are often intentionally employ some measures for evasion (Maor, 2013).

Unlike the surveys discussed above, Abdel-Hafez and Xu (2013) discuss and compare existing user modelling techniques for social media sites. They also explain how user profiles are constructed in their modelling process. Jin et al. (2013) review user behaviours in online social networks by social connectivity, interaction among users, and traffic activity. They also analyse malicious behaviours of online social network users and proposed solutions to detect misbehaving users. The focus is, however, on user social behaviours rather than security. Stamatatos (2009) surveys automated approaches to attributing authorship by examining their profiles for both text representation and text classification. However, the focus of this survey is on computational requirements and settings rather than on linguistic or literary issues. Rodríguez et al. (2014) classify human activity recognition methods as data-driven and knowledge-based techniques and use them to represent human