JOURNAL OF INFORMATION SECURITY AND APPLICATIONS 32 (2017) 27–46

Available online at www.sciencedirect.com

ScienceDirect

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

A white-box anomaly-based framework for database leakage detection

E. Costante [a,*], J. den Hartog [a], Milan Petković [a,b], S. Etalle [a,c], M. Pechenizkiy [a]

[a] Eindhoven University of Technology, The Netherlands
[b] Philips Research Europe, High Tech Campus, The Netherlands
[c] University of Twente, The Netherlands

ARTICLE INFO

Article history:
Available online 15 November 2016

Keywords:
Data leakage
Insider threats
Database monitoring
Database intrusion detection

ABSTRACT

Data leakage is at the heart most of the privacy breaches worldwide. In this paper we present a white-box approach to detect potential data leakage by spotting anomalies in database transactions. We refer to our solution as white-box because it builds self explanatory profiles that are easy to understand and update, as opposite to black-box systems which create profiles hard to interpret and maintain (e.g., neural networks). In this paper we introduce our approach and we demonstrate that it is a major leap forward w.r.t. previous work on the topic in several aspects: (i) it significantly decreases the number of false positives, which is orders of magnitude lower than in state-of-the-art comparable approaches (we demonstrate this using an experimental dataset consisting of millions of real enterprise transactions); (ii) it creates profiles that are easy to understand and update, and therefore it provides an explanation of the origins of an anomaly; (iii) it allows the introduction of a feedback mechanism that makes possible for the system to improve based on its own mistakes; and (iv) feature aggregation and transaction flow analysis allow the system to detect threats which span over multiple features and multiple transactions.

1. Introduction

Data represent a great value for most organizations. Databases, storing customer and confidential business data, are a core asset that needs to be protected from illegitimate usage. This makes data leakage, i.e. the unauthorized/unwanted transmission of data and information (Gordon, 2007), a major threat (Opderbeck, 2016). According to a Ponemon Institute study (http://www.ponemon.org/news-2/23), in 2009 data breach incidents cost U.S. companies an average of $6.75 million per-incident. These costs include theft of corporate intellectual property, damages to reputation and decrease in costumers' trust. To reduce these enormous costs and comply with legislation, a timely detection of data leakage is essential. Especially, the new EU Data Protection regulation requires a prompt notification of a data breach to the data subject (Information Age, 2013): in this case, the white-box nature of the approach we propose helps identifying the exact nature of the breach and the individuals involved.

* *Corresponding author.* Eindhoven University of Technology, The Netherlands.
E-mail address: e.costante@tue.nl (E. Costante).
http://dx.doi.org/10.1016/j.jisa.2016.10.001

Insider threats, i.e. careless, unhappy or malevolent employees, are amongst the main sources of data leakage. Since insiders have the right of accessing internal resources such as databases, they can harm a company more easily than external threats such as hackers. Leakages from database account for most of the records globally disclosed in 2012 (Verizon, 2013). For these reasons, we focus our work on the problem of detecting database leakages caused by insiders.

A first line of defence against data leakage is formed by Access Control (AC) mechanisms (Samarati and de Vimercati, 2001) which aim to regulate users' rights to access certain data. AC suffers of some disadvantages, e.g. it is not always expressive enough to allow the definition of fine-grain access rules and, especially in dynamic domains, it has high costs of maintainability (update of rights and roles). In addition, AC might reduce data availability which is critical in emergency situations (e.g. in healthcare domains) or productivity (e.g. time loss to ask for permission to access certain documents). As a result, organizations often apply relaxed AC policies which give users access to more information than they actually need (Caputo et al., 2009). Besides AC, there are tools and methodologies for data leakage detection (Shabtai et al., 2012) which can spot leakages by operating at different locations on the data path (e.g. network, workstation or database). In this paper we focus on solutions operating at a database level so that leakages can be detected at a very early stage, i.e. when sensitive data is leaving its primary source.

Commercial tools and academic solutions addressing database leakage detection typically work by monitoring database traffic in terms of SQL queries. Existing solutions can be divided into *signature-based* and *behavioural-based* systems (Patcha and Park, 2007). Generally, in signature-based systems a blacklist defines the set of dangerous or denied access patterns. On the other hand, behavioural-based systems automatically learn permitted access patterns by observing normal activities and mark every anomaly as a potential threat. The main problem of signature-based approaches is that they can only detect well-known attacks, whereas behavioural-based approaches have the great potential of detecting unknown database attacks. In addition, by automatically generating fine-grained profiles, behavioural-based solutions reduce the administrative burden for setting up the system thus lowering costs.

These advantages make behavioural-based approaches widely adopted in literature (Bockermann et al., 2009; Chung et al., 2000; Fonseca et al., 2007; Gafny et al., 2011; Kamra et al., 2008; Mathew and Petropoulos, 2010; Roichman and Gudes, 2008; Santos et al., 2012; Wu et al., 2009). However, these approaches have certain drawbacks. The first problem is the high *False Positive Rate* (FPR) they usually generate. Since each false alert has to be analyzed by a security officer, false positives have a high operational cost. In network anomaly detection (Bolzoni et al., 2009; Hadžiosmanović et al., 2012) (a different yet related field), a system starts to be "usable in practice" when it shows a FPR in the order of 0,01%, a rate by far not attained by present database anomaly detection systems. A way to keep FPR low is to frequently update normal usage profiles, so that they reliably represent actual normal behaviour. However, profiles update is usually a costly operation since it often requires re-training the model. The second drawback is that current solutions provide little or no support for alert handling. Usually, when an alert is raised, it is accompanied by a *deviation degree*, or an *anomaly score*. Unfortunately, this information is virtually useless for the security officer as it does not support him in understanding "what is going on". To this end, signature-based systems have an advantage: when they raise an alert, they can say exactly which policy is violated and why this violation may constitute a problem. For behaviour-based systems, it is more difficult to "explain" the reasons of an anomaly mainly because of their *black-box* nature, i.e. the underlying engine (be it a neural network or a machine learning classifier) is difficult to understand by a human.

To enable practical detection of unknown database leakage threats, this work extends the work done in Costante et al. (2014) which, to the best of our knowledge, is the first *white-box* behavioural-based database leakage detection system. The white-box approach has also proven successful in other domains such as intrusion detection for industrial control systems (Yüksel et al., 2016a) and within back-office networks (Yüksel et al., 2016b). It can also be combined with preventative measures (Costante et al., 2016) and enables further investigation of the anomalies (Vavilis et al., 2015). Our system advances the state-of-the-art in the following ways:

1. We propose a comprehensive database leakage detection framework. We build histogram-based profiles over a wide feature space which leads to the construction of fine-grain profiles and allows the detection of several types of attacks. Thanks to the white-box approach, users can easily understand what profiles mean in terms of database activities, and manually inspect and refine profiles if necessary;
2. When an alarm is raised, we clearly determine the origins of each anomaly, hence facilitating the security officer's handling of the alarms;
3. The use of histogram-based profiles enables online-learning (profiles are incrementally built) and facilitates updates to the model without requiring the costly re-training of the complete model. Adding a new sample to the model will only impact the frequencies of the associated bin, which has very limited computational costs;
4. We introduce a feedback mechanism which enables the security officer to mark false positives. This mechanism has the advantage of: (i) speeding-up the post-processing of alarms; and (ii) allowing the update of existing profiles to progressively reduce FPR;
5. We introduce a new mechanism to aggregate features based on a coupling score which identifies pairs of features that help to detect more threats if considered together rather than independently;
6. We propose a transaction flow analysis which enables the detection of attacks spanned over multiple transactions;
7. We evaluate the performance of our framework with an extensive set of experiments carried out over two different datasets, one created from simulated scenarios, and the other consisting of more than 12 millions of real transactions coming from an enterprise operational database. In addition, in our experiments we benchmark our system against other approaches from the literature.

The remainder of this paper is structured as follows: in Section 2 we describe the state-of-the-art of this field, while