JOURNAL OF INFORMATION SECURITY AND APPLICATIONS 29 (2016) 80–92

Available online at www.sciencedirect.com

ScienceDirect

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

Enabling information recovery with ownership using robust multiple watermarks

Vidhi Khanduja [a,*], *Shampa Chakraverty* [a], *Om Prakash Verma* [b]

[a] *Department of Computer Engineering, Netaji Subhas Institute of Technology, Delhi, India*
[b] *Department of Computer Science and Engineering, Delhi Technological University, India*

ARTICLE INFO

Article history:
Available online 8 April 2016

Keywords:
Information recovery
Database watermarking
Right protection
Robustness
Tamper detection

ABSTRACT

With the increasing use of databases, there is an abundant opportunity to investigate new watermarking techniques that cater to the requirements for emerging applications. A major challenge that needs to be tackled is to recover crucial information that may be lost accidentally or due to malicious attacks on a database that represents asset and needs protection. In this paper, we elucidate a scheme for robust watermarking with multiple watermarks that resolve the twin issues of ownership and recovery of information in case of data loss. To resolve ownership conflicts watermark is prepared securely and then embedded into the secretly selected positions of the database. Other watermark encapsulates granular information on user-specified crucial attributes in a manner such that the perturbed or lost data can be regenerated conveniently later. Theoretical analysis proves that the probability of identifying target locations, False hit rate and False miss rate is negligible. We have experimentally verified that the proposed technique is robust enough to extract the watermark accurately even after 100% tuple addition or alteration and after 98% tuple deletion. Experiments on information recovery reveal that successful regeneration of tampered/lost data improves with the increase in the number of candidate attributes for embedding the watermark.

1. Introduction

With the growth of Information and Communication Technology in the progress of mankind, the information has taken center-stage. A major proportion of the internet content is dynamically generated from databases. This has driven the capabilities, sizes and performance of databases to grow in exponential magnitudes. In this scenario where end users are demanding more and more information to be available on the net, data providers are burdened to supply accurate data and at the same time ensure its security. A major threat faced is that unauthorized and illicit copies of true data can be easily generated and distributed using the same enabling technologies. To counter such attacks, legal as well as technological solutions are being devised to assert data ownership. Watermarking is one such widely accepted technological measure to protect the data (Kerr et al., 2003).

* *Corresponding author.* Department of Computer Engineering, Netaji Subhas Institute of Technology, Delhi, India. Tel.: +919999695797.
E-mail address: vidhikhanduja9@gmail.com (V. Khanduja).
http://dx.doi.org/10.1016/j.jisa.2016.03.005

High risk databases such as databases from medicine, weather, transportation, automobile, military, etc., contain critical attributes. The information conveyed by critical attributes needs to be preserved. Additionally, these high-risk databases are subject to various malicious attacks resulting in threat to their ownership (Agrawal et al., 2003).

This motivated us to propose a technique that not only recovers information irrespective of a number of alterations occurred but also resolves ownership issues. We follow an information-centric data recovery technique to preserve the information. The proposed robust watermarking technique prepares dual watermarks and securely conceals them within the database. One watermark contains the ownership details while the second one holds information to be preserved.

This paper proceeds with related work in Section 2. Section 3 elucidates the proposed watermarking technique. Section 4 presents the Security Analysis of the proposed scheme. Section 5 presents experimental results and analysis to verify robustness and information recovery followed by comparative evaluation of our work with a prior work. We conclude the proposed technique in Section 6.

2. Related work

Literature is rife with interesting works in the domain of robust watermarking for digital databases. R. Agrawal, Peter J. Haas, J. Kiernan identified the need of database watermarking in 2002 (Agrawal et al., 2003). They proposed that database relations can be watermarked in some algorithmically selected attributes out of several candidates attributes in a tuple. The technique suffered from certain drawbacks. The technique is primary key dependent and therefore is not suitable for databases that does not contain primary key. Also, technique does not provide mechanism for multi-bit watermark. Their technique generates limited potential locations that can be used to hide watermark bits without being subjected to removal or destruction. Several techniques proposed in literature enhanced the work of Agrawal et al. (2003) by embedding multi-bit watermark in selected LSBs (Ali and Mahdi, 2011; Cui et al., 2007; Farfoura et al., 2012).

Xinchun, Xiaolin and Gang have proposed weighted algorithm (Cui et al., 2007). The technique assigns the weight to attributes according to their significance. This result in the increase in the chances of selection of high rank significant attributes as compared to less important ones. The mark bit-string comes from meaningful character of a database owner such as names. Robustness against subset deletion, addition and alteration attacks is discussed. However, primary key dependent embedding process is based on work proposed in Agrawal et al. (2003). In Ali and Mahdi (2011), Dr. Yossra H. Ali and Bashar Saadoon Mahdi proposed a technique that uses threshold generator based on simple combination of odd number of register. However, this technique has same limitations as in Agrawal et al. (2003).

In Farfoura et al. (2012), they proposed time-stamping based protocol to resolve additive attacks. The embedding process was reversible so that original values can be regained after extraction. Reversibility allows one to recover the original data completely from the watermarked database after authenticating with a time-stamp protocol. However, technique can be applied where data authentication and original content recovery were required at the same time. Once the original data is recovered, the database loses ownership protection. Recently, semi-blind reversible technique has been proposed in Ifthikar et al. (2015). In this technique, the knowledge of mutual information for every candidate feature is employed to create watermark. Genetic algorithm is used to resolve the constraint optimized problem of optimal watermark encoding to ensure high data quality.

Most of the prior works on robust watermarking assume the watermark to be a random bit stream, which is otherwise meaningless (Agrawal et al., 2003; Ali and Mahdi, 2011; Cui et al., 2007; Shehab et al., 2008; Sion et al., 2004). Little other notorious work based on fragile watermarking schemes exists in literature (Camara et al., 2014; Guo et al., 2006; Guo, 2011; Khan and Husain, 2013; Li et al., 2004), resolving tamper detection and localization of perturbation issues. However, all these works do not take into account the Data/Information recovery. The application of watermarking in data recovery has received scant attention so far.

To our knowledge, the only work that reports a watermarking scheme for recovering data is by Khataeimaragheh and Rashidi (2010). In this work, authors have proposed a fragile watermarking technique that can detect and correct distortions in RDBs by embedding watermarks created from each attribute value, thereby recovering true data. There are certain serious shortcomings in their approach. Firstly, it can only be used to detect and recover altered data and preclude proper recovery from deletions. Secondly, the probability of accurately detecting, localizing and hence rectifying errors reduces drastically when the number of errors exceeds two.

In contrast to prior work, our proposed approach deciphers data in terms of information it represents and helps recover it from altered as well as deleted data. We utilize the watermark as information carrier to address the need to recover lost information due to frequent attacks. We propose a scheme for robust watermarking with multiple watermarks that resolve the twin issues of ownership along with recovery of information.

3. Architecture of the proposed scheme

Information is framed in the form of a relational database $\mathcal{R}$. The database $\mathcal{R}$ is built upon a relational model. We define relational schema as $\mathcal{R}(\mathcal{K}_p, \mathbf{A})$ comprising N_t tuples, i.e. $N_t = |\mathcal{R}|$. $\mathcal{K}_p$ is the primary key attribute and $\mathbf{A}$ is a set of attributes such that $\mathbf{A} = \{A_1, \ldots, A_{N_a}\}, |\mathbf{A}| = N_a$. $\mathbf{A}$ represents a set of attributes in $\mathcal{R}$ excluding $\mathcal{K}_p$.

Fig. 1 depicts block diagram of the proposed watermarking scheme showing the main components. The proposed scheme consists of four phases: Watermark Preparation, Watermark Insertion, Watermark Extraction, and Decision