Journal of Information Security and Applications 37 (2017) 38–49

Contents lists available at ScienceDirect

Journal of Information Security and Applications

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

Database watermarking, a technological protective measure: Perspective, security analysis and future directions

Vidhi Khanduja

Department of Computer Engineering, NSIT, Delhi, India

ARTICLE INFO

Article history:

Keywords:
Information security
Digital watermarking
Right protection
Tamper detection

ABSTRACT

Digital Databases dynamically generates a major proportion of the internet content. The databases are created, stored and accessed digitally and transmitted through computer networks. This has grown the potential, sizes and performance of databases in exponential magnitudes. Thus, the need to protect digital databases arises due to the increased vulnerability to copyright and piracy threats originating from the Internet. Both legal and technological measures must be utilized in a synergetic manner to ensure an adequate level of protection. TPMs backed by legal anti-circumvention measures offer a cost-effective solution to database protection. We provide the current state-of-art and analyses of the arena of digital database protection from a combined legal and technical perspective. Our work is more focused on security analysis of the work done so far and provides readers with detailed discussion on the future directions in the domain of digital watermarking of databases.

1. Introduction

Databases are repertoires of knowledge garnered by the collective efforts of mankind through ages and across regions. Digital Databases dynamically generates a major proportion of the internet content. The databases are created, stored and accessed digitally and transmitted through computer networks. This has grown the potential, sizes and performance of databases in exponential magnitudes. Whether they are sold in pieces for data mining applications such as stock market data, consumer behaviour data, power consumption data and weather data or maintained in-house such as product data by e-commerce sites and medical history of patients by hospitals, databases play a pivotal role in all aspects of society.

As end users demand more and more information to be available on the net either at low or no cost, database developers are interested in generating revenues from creating databases as this involves intellectual and financial inputs. The need to protect databases arises due to the increased vulnerability to copyright and piracy threats originating from the Internet [1]. Developers have the responsibility to not only supply accurate data, but also ensure its security against illegal copying, hacking or tampering. The digital watermarking of copyrighted works augmented by legal means is fast emerging as an effective and efficient means to protect shared and outsourced databases from infringement.

Section 2 introduces the process of digital watermarking in databases. In Section 3, we analyse the arena of digital database protection from a combined legal and technical perspective. Section 4 throws the light on existing work in literature; Section 5 presents the security analyses of the various techniques and Section 6 discusses the future directions. Finally, Section 7 concludes the paper.

2. An overview of digital database watermarking

Digital watermarking is a viable and cost-effective technological method that protects digital documents such as images, video and databases by *marking* them with some digital pattern. Watermarking algorithms for digital databases invariably introduce small changes in the data being watermarked with an objective of inserting the mark, but without altering the database in *any significant way*. Watermarking does not completely prevent piracy. However, it does provide a means to establish the true identity of the owner and deters attempts to plagiarize or distort it [2]. Fig. 1 illustrates the process of watermarking databases. The watermark is secretly prepared/selected and concealed within the database. The process of embedding the watermark is made secure by using secret parameters. The watermarked database armed with watermark is now available on web for its intended application. In case of any ownership disputes or database is tampered with, the watermark is extracted from the suspected database. The robust watermarking techniques resolves ownership issues while fragile watermarking is for integrity violation.

Unlike encryption and hashing techniques, watermarking does not attempt to hide the data, but instead infuses a kind of ownership proof in the data. Encryption and hashing provide protection

https://doi.org/10.1016/j.jisa.2017.10.001
2214-2126/© 2017 Elsevier Ltd. All rights reserved.

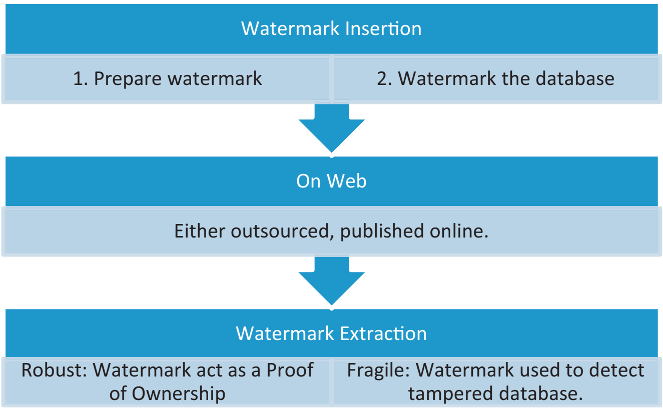
Watermark Insertion
1. Prepare watermark
2. Watermark the database
On Web
Either outsourced, published online.
Watermark Extraction
Robust: Watermark act as a Proof of Ownership
Fragile: Watermark used to detect tampered database.

Fig. 1. Watermarking relational databases.

to the content by making the information indecipherable by an attacker. On the contrary, digital watermarking works on the principle of modifying the content in a manner so that the usability of the data is retained fully. In fact, an attacker or observer has no way to decipher that a watermark is actually present in the database. However, the watermark does remain within the content inseparably providing a proof of the ownership or a signature to detect tampering or to track the people who may have obtained the content legally and are illegally redistributing it [3].

Many researchers have contributed significantly in the area of watermarking multimedia data such as image, audio and video sources [3–7]. The technique to conceal watermark within databases differs significantly with that of multimedia data [2]. As compared with data in relational databases, multimedia data is voluminous and has a large bandwidth to hide watermarks in a redundant manner. Moreover, the relative spatial positioning of different parts of multimedia data such as image or video is not affected significantly when watermark bits are inserted. Therefore, the quality of image is largely retained.

Relational databases on the other hand, comprise concrete tuples and attributes, each tuple representing a distinct entity. Firstly, we need to disperse the watermark bits across multiple tuples to achieve redundancy. There is no particular ordering between these tuples, so even a subset of tuples can be used. Secondly, embedding of watermark bits invariably introduces perturbations within a database. These perturbations may adversely affect the usability of the database. Therefore, a database watermarking technique must ensure that the usability constraints of the attributes stay within limits when watermark bits are inserted. Usability constraints are the limitations imposed on each attribute. They are decided by the database owner or designer and depend upon its specific application(s). For example, attribute value should be unique; classification range must remain same before and after concealing watermark, etc. [8]. Any watermarking model for databases must satisfy certain properties. We enumerate the important properties below [2]:

A. **Imperceptibility:** The amount of perturbations made to data must be such that its usability is still maintained. This property is referred to as imperceptibility. Encyclopaedia Britannica has embedded watermarking by making small changes to population and surface area of countries [2]. Furthermore, weather data can tolerate the error in daily temperatures of 1° or 2° [2]. These usability constraints give a hint of where a watermark should be concealed conveniently within the database. Similarly, the publishers of books of mathematical tables such as logarithm tables and astronomical ephemerides have introduced small errors in their tables through centuries, so as to identify pirated copies [2].

B. **Blind watermarking**: A watermarking technique is blind if it does not require the original un-watermarked database for watermark extraction. Blind watermarking is also referred to as oblivious watermarking.

C. **Embedding effectiveness**: The watermarking technique should be such that it can be applied to successfully embed a watermark in any randomly selected database. The technique should not be specific to any particular database.

D. **Robustness:** This is the ability of the watermark to resist perturbations caused by benign or malicious attacks. For most applications that mandate copy-control, robustness is the major concern. The watermark must firmly establish the ownership of the database. Such watermarks should be extracted even after several random alterations caused by various attacks.

E. **Security:** According to Kerckhoff's law, watermarking algorithms are publicly known to everyone [2]. Its security lies in the cryptographic key. A watermarking technique is secure if an attacker is not able to detect or remove the watermark even after knowing the algorithms for embedding and extraction.

F. **Low-complexity:** In any watermark scheme, the insertion, detection and extraction processes must employ efficient algorithms with low time complexity.

G. **False Hit Rate**: It is the probability of detecting our watermark in someone else's non watermarked relation. False Hit Rate should be low for a watermarking system.

H. **Incremental updatability**: Watermarks should be incrementally updatable, i.e., as the attributes of the tuples in a database are altered (added, modified or deleted), the watermark should be recomputed for only such modified tuples.

3. Database watermarking as a technological protection measure with legal protection

The foundation of modern society is intelligible information compiled in innumerable databases. Government departments, corporations, multinational companies, information bureaus and research centers produce databases of government records, medical and legal case records, web pages and collection of literary and artistic works.

Article 1 (2) of Directive 96/9/EC of the European Parliament defines a database as "*a collection of independent works, data or other materials arranged in a systematic and methodical way and individually accessible by electronic or other means*" [9]. According to the Black's Law Dictionary, a database is defined as "*a compilation of information arranged in a systematic way and offering a means of finding specific elements it contains, often today by electronic means*" [10]. The relevance of digital databases is explicitly highlighted in this definition. The distinguishing feature of digital databases is that they can be created, downloaded, value added and digitally re-transmitted with great flexibility and speed.

It requires significant efforts in terms of money, manpower and creative inputs to build high-quality databases. They are thus Intellectual Property in their own right. Given their rich informational content and the ready availability of advanced technologies to communicate and modify them with relative ease, it is imperative to protect digital databases against potential misuse. Technological methods that are designed to protect digital content of any form like text, images and databases are called "Technological Protection Measures (TPM). They include the technologies that can control access to copyrighted digital content or can prevent users from copying such protected content. Watermarking of digital databases is one such TPM that has emerged as an effective