Effective Approaches for Watermarking XML Data

Wilfred Ng and Ho-Lam Lau

Department of Computer Science,
The Hong Kong University of Science and Technology, Hong Kong
{wilfred, lauhl}@cs.ust.hk

Abstract. Watermarking enables provable rights over content, which has been successfully applied in multimedia applications. However, it is not trivial to apply the known effective watermarking schemes to XML data, since noisy data may not be acceptable due to its structures and node extents. In this paper, we present two different watermarking schemes on XML data: the selective approach and the compression approach. The former allows us to embed non-destructive hidden information content over XML data. The latter takes verbosity and the need in updating XML data in real life into account. We conduct experiments on the efficiency and robustness of both approaches against different forms of attack, which shows that our proposed watermarking schemes are reasonably efficient and effective.

1 Introduction

Watermarking in the contexts of image, audio or video data is well-known to be an effective technique to protect the intellectual property of electronic content. Essentially, the technique embeds a secret message into a cover message within the content in order to prove the ownership of materials. Remarkable successes in watermarking on multimedia applications have been achieved in recent years [4]. Thus, relevant business sectors are able to distribute their data while keeping the ownership and preventing the original data being resold illegally by others.

The existing watermarking technology has mostly been developed in the context of multimedia data, since such data has a high tolerance to noise and thus it is not easy to detect the watermark. Unlike multimedia data, XML data are diverse in nature: some are data-centric and numeric (e.g. regular scientific data) while some are document-centric and verbose (e.g.book chapters). It is challenging to develop an effective watermarking scheme which is invisible and is able to resist various kinds of attack.

In this paper, we attempt to develop watermarking schemes for XML data based on two different watermarking approaches. One is the *selective approach* and another is the *compression approach*. As for the selective approach, we develop a watermarking scheme for uncompressed XML data based on the database watermarking algorithm proposed by Agrawal [2]. The second approach is more interesting. It follows our advocation that in reality some XML documents are verbose and they need compression in practical applications [3]. In addition, we

L. Zhou, B.C. Ooi, and X. Meng (Eds.): DASFAA 2005, LNCS 3453, pp. 68–80, 2005.

take into consideration that XML documents need to be updated frequently. Therefore, in the compression approach, we introduce a novel watermarking scheme based on our earlier developed XML compressor, namely XQzip, which does not require full decompression when querying. [3]. By watermarking compressed XML data, we gain the advantage of having better document security, and at the same time, higher flexibility of updating XML data.

Related Work. Agrawal presents an effective watermarking technique for the relational data [2]. This technique ensures some bit positions of certain attributes contain the watermarks. We extend their techniques on XML data by defining locators in XML in our selective approach. Sion [5] discusses the watermarking of semi-structures of multiple types of contents and represents them as graphs by characterizing the values in the structure and individual nodes. He also proposes a watermarking algorithm that makes use of the encoding capacity of different types of nodes. Gross-Amblard [1] investigates the problem of watermarking XML databases while preserving a set of parametric queries. His work mainly focuses on performing queries on different structures and pay less attention to the watermarking scheme. However, the query approaches are similar to the predefined queries used in the compression approach. At present, all proposed XML watermarking schemes are based on uncompressed XML data and no studies exist on watermarking compressed XML data to the best of our knowledge.

Paper Outline. After introducing our XML watermarking schemes in this section, we describe and study the selective approach, which is for uncompressed XML data, and the compression approach, which is for XQzip compressed XML data in Sections 2 and 3, respectively. Then in Section 4, we conclude our work and suggest future improvements for the watermarking schemes we developed.

2 The Selective Approach of Watermarking XML

In this section, we introduce the selective approach of XML watermarking. We also analyze the robustness of our watermarking system against the following two forms of attacks: *subtractive attack* and *additive attack*. All the experiments related to the watermarking system are conducted on a machine of the configuration as follows: P4 2.26GHz, 512MB main memory and 15GB disk space.

2.1 Watermark Insertion

In the selective approach of watermarking XML, the watermarks are randomly distributed throughout the XML document based on a secret key provided by the owner. We aim at making minor changes on XML data without causing errors during the process.

The watermark insertion algorithm and the notations we used are presented in Algorithm 1 and Table 1, respectively. Before embedding marks in XML data, we define a *locator*, which is an analogy to the primary key in relational databases, to indicate whether a particular element should be marked. Unlike