Query-Preserving Watermarking of Relational Databases and XML Documents

DAVID GROSS-AMBLARD
Laboratoire Le2i UMR CNRS 5158

Watermarking allows robust and unobtrusive insertion of information in a digital document. During the last few years, techniques have been proposed for watermarking relational databases or XML documents, where information insertion must preserve a specific measure on data (for example the mean and variance of numerical attributes).

In this paper we investigate the problem of watermarking databases or XML while preserving a set of parametric queries in a specified language, up to an acceptable distortion. We first show that unrestricted databases can not be watermarked while preserving trivial parametric queries. We then exhibit query languages and classes of structures that allow guaranteed watermarking capacity, namely 1) local query languages on structures with bounded degree Gaifman graph, and 2) monadic second-order queries on trees or tree-like structures. We relate these results to an important topic in computational learning theory, the VC-dimension. We finally consider incremental aspects of query-preserving watermarking.

Categories and Subject Descriptors: H.1 [Information Systems]: Models and Principles; F.1.3 [Theory of Computation]: Complexity Measures and Classes; F.4 [Theory of Computation]: Mathematical Logic and Formal Languages

General Terms: Theory, Security, Algorithms

Additional Key Words and Phrases: VC-dimension, watermarking

1. INTRODUCTION

Context. A growing part of Internet content is dynamically generated from databases. Classical situations can be captured by the following 3-tier model: data owners invest time and efforts to elaborate large and detailed databases, and wish to sell them to multiple data servers. Data servers buy such data and answer queries, for example through a Web interface, to several final users. A simple example is given by air travel information websites: timetables of all flights are possessed by a data owner. Data servers propose a search engine on these flights, answering queries asked by final users such as “flights from Paris to Delhi on the 1st of July”. Other common examples are lodging information systems, meteo-
logical and financial data, etc. But data owners are exposed to malicious servers, trying to sell illegal copies as their own. This problem is strengthened by the digital nature of this information, since perfect copies of a data set are easily produced and disseminated. Thus, an important tool for data owners is the ability to argue ownership of a database, once a suspect one has been discovered, and to track back to the original malicious server.

Database watermarking. Informally, database watermarking techniques hide pertinent information in a data set, such as the owner or purchaser’s identity. This information is used to identify the real owner of a suspect data set, or the original server who has performed an unauthorized diffusion. During the last few years, algorithms were proposed for watermarking structured data like XML documents and relational databases (for example [Agrawal et al. 2003; Gross-Amblard 2003; Sion et al. 2004; Li et al. 2005; Lafaye et al. 2008]). All these methods hide information by a controlled alteration of data. For example, Agrawal et al. [2003] flip secret bits in numerical attributes, and limit alteration by focusing on least-significant bits. Based on experimental results, they observe that precision, mean and variance of numerical attributes are hardly impacted by this operation. But, as the majority of the other proposed database watermarking methods, there is no guarantee for the distortion induced on query answers.

Query-preserving watermarking. In this paper we focus on the watermarking of databases, in the general setting where data servers perform queries in a language $L$. The data owner has a valuable database instance, and data servers apply for a copy of this database, providing queries $\psi_1, \ldots, \psi_k$ they will answer to final users. These queries will be parametrized by final users’ inputs. The problem is then to construct a query-preserving watermarking protocol that respects the following conditions:

— the protocol hides a message $m$ into the owner’s database instance, and induces a guaranteed distortion on the results of queries $\psi_1(\bar{a}), \ldots, \psi_k(\bar{a})$, for any user input $\bar{a}$.

— the protocol is able to extract the message hidden in a suspect data set, based on answers to queries $\psi_1, \ldots, \psi_k$ only (the owner acts as any final user to get these answers).

— the protocol fails extracting messages from unwatermarked databases.

— the protocol is robust against reasonable attacks, like random distortion of query answers.

— the protocol has a scalable capacity: the amount of hidden bits grows with the database size.

Such a watermarking protocol is driven by what is important to the final user: results of queries. Notice that data servers may answer other queries to users, but only distortion on $\psi_1, \ldots, \psi_k$ is guaranteed.

Example 1.1. Tables below present a touristic database instance. The goal is to hide information by slightly modifying transport prices so that the total cost of any travel remains unchanged. This is equivalent to preserving the following aggregate query, for any value of parameter $t$:

ACM Transactions on Database Systems, Vol. V, No. N, Month 20YY.
\[ \psi(t) \equiv \text{select sum(price) from PriceTable, Route} \]

\[ \text{where PriceTable.transport = Route.transport} \]

\[ \text{and travel} = t. \]

**Watermarking capacity.** The fundamental difficulty in query-preserving watermarking is to obtain a protocol with a reasonable bandwidth. For example, if \( n \) bits can be hidden in a data set, \( 2^n \) distinct messages can be hidden. By mapping distinct messages with distinct data servers, the message extraction allows to identify the suspect server that performed a copy. Another application is to combine the hidden message with an error-correcting code in order to increase its robustness against attacks. Hence, obtaining lower bounds on the watermarking capacity is interesting.

In this direction, Khanna and Zane [2000] proposed a protocol with guaranteed capacity for a specific parametric query: shortest path queries on weighted graphs. Their information insertion does not modify the length of shortest path between any pair of vertices beyond an acceptable distortion. The watermarking capacity is \( \Omega \left( \frac{1}{n^{1/2 - 1/d}} \right) \) for distortion \( d \), for any graph with \( n \) edges. From the theoretical point of view, they observe that shortest path queries have a low computational complexity, and suspect that watermarking protocols for NP-hard search spaces are difficult to analyze.

**Contribution:** watermarking and learning theory. In this paper, we study the capacity of watermarking protocols that preserve any set of predefined queries from a language \( \mathcal{L} \). Up to our knowledge, this is the first work dealing with this problem in the database theory perspective. Our goal is not to propose a new watermarking algorithm for databases, but to study the potential watermarking bandwidth that can be awaited from given data sets and queries (the present work has nevertheless a practical counterpart [Lafaye et al. 2008] as explained in the related work).

Our first contribution is an extension of Khanna and Zane [2000]'s graph watermarking model into the relational setting. This natural extension requires some care, as a general query on a general structure is much more intricate than properties considered in their work (shortest paths on graphs, which are simply subsets of the graph). We then study the complexity of computing the exact watermarking capacity, which turns out to be intractable.

Our first main result shows that the difficulty of query-preserving watermarking is linked to the informational complexity of sets defined by queries, rather than their computational complexity. This is related to an important combinatorial
parameter in computational learning theory, the Vapnik-Chervonenkis dimension of sets (or VC-dimension, see for example [Blumer et al. 1989]). Roughly speaking, a set is said to be shattered if all its subsets can be defined by a query, and the VC-dimension represents the largest cardinality of such shattered sets. It is well known that a finite VC-dimension for a family of sets is equivalent to its learnability (in the PAC learning model [Valiant 1984]). In summary, our result states that if the VC-dimension is not bounded but is maximal on all input structures, that is, if any set is shattered by queries, no watermarking protocol can be obtained.

Grohe and Turán [2004] showed that the VC-dimension of sets defined by first-order logic and monadic second-order logic is bounded on restricted classes of structures, and this characterization is, in some sense, optimal. These restrictions concern bounding the degree of the Gaifman graph of the structure, or bounding its tree-width, which measures its similarity with trees. This latter restriction has also fruitful applications in both database theory and computational complexity (see for example [Flum et al. 2002]).

Our second main result shows that under the same restrictions, a watermarking protocol with guaranteed capacity can be obtained. First, we construct a watermarking protocol for database instances with bounded degree Gaifman graph, preserving any local query. Local languages contain particularly first-order logic, order-invariant queries [Grohe and Schwentick 2000], and relational AGGREGATE queries [Hella et al. 2001], that expresses mostly plain SQL by adding grouping and aggregate functions to relational calculus [Grumbach et al. 1996; Libkin and Wong 1997]. Second, we provide a watermarking protocol for first-order and monadic second-order queries on trees or tree-like structures. Monadic second-order logic (MSO) is of a special interest, since it is commonly used to model pattern queries on labeled trees, as a formal query language for XML documents (see for example [Neven and Schwentick 2002]).

Finally, on structures with unbounded degree Gaifman graph, one can construct a first-order formula that defines sets with unbounded and maximal VC dimension. There also exists an MSO-formula yielding such sets on structures with unbounded tree-width. For both, no query-preserving watermarking protocol can be obtained. This maps a rather complete panorama of query-preserving watermarking.

Organization. The paper is organized as follows: we first give basic definitions on query-preserving watermarking on Section 2. We show on Section 3 that computing the exact watermarking capacity is hard, and that no watermarking protocol can be obtained for unrestricted database instances, even for trivial queries. We then exhibit restrictions on database instances and query languages that allow watermarking with a reasonable amount of hidden information: local languages on structures with bounded degree Gaifman graph on Section 4, monadic second-order queries on trees and tree-like structures on Section 5. Section 6 extends the previous results so that malevolent attacks are taken into account. Complementary results on easier database watermarking scenarios are given in Section 7, and the incremental updatability of watermarked instances is exposed in Section 8. Finally, Section 9 compares the present method with its practical counterpart, and the last section concludes.
Related work. A wide part of the watermarking literature focuses on multimedia data including images, sound and video [Zeng et al. 2006; Cox et al. 2001; Katzenbeisser and Petitcolas 2000]. The watermarking capacity is a classical topic in multimedia watermarking, where the common tool is information theory [Cappe\-lini et al. 2001]. But this community does not consider intricate query answers on structured data sets, but rather global metrics like peak signal-to-noise ratio (PSNR) on a signal (an image) without formal structure.

Beside works cited in the introduction [Agrawal et al. 2003; Khanna and Zane 2000], watermarking of structured data like trees, graphs, or solutions of an optimization problem are studied in [Qu and Potkonjak 1999; 2003; Wong et al. 2004; Wolfe et al. 2001]. These approaches do not consider the notion of queries. Sion et al. [2004] and Gross-Amblard [2003] introduced query-preservation methods for watermarking. Sion et al. [2004] handle potentially any kind of constraints by a greedy procedure calling external checking programs (usability plugins). Shehab et al. [2008] extend this work so that better watermarks are obtained. These very general methods do not elaborate on the syntactical form of queries in order to find valid watermarks. Gross-Amblard [2003] and Lafaye et al. [2008] take into account the syntactical form of aggregate constraints, so that the watermark search procedure is optimized.

Our formalism follows the approach of Khanna and Zane [2000]. In this work, a unique parametric query is considered, namely shortest paths in weighted graphs. This formal approach was first generalized in [Gross-Amblard 2003]. In this latter work, a lower bound on the watermarking capacity is obtained for queries taken from several query languages, along with an upper bound on data distortion. No lower bound on data distortion was provided. In the present work, we propose such a lower bound. We also provide a watermarking protocol achieving this minimal error. Hence this bound is tight. Finally, even without a malevolent attacker, the previous method was probabilistic, while the method presented in this paper is deterministic. We also provide results on different scenarios, where the data owner has full access to the suspect instance, and when non-parametric queries are to be preserved.

2. QUERY-PRESERVING WATERMARKING

2.1 Basic definitions

Weighted structures. In this paper, attention is focused on watermarking numerical data for the sake of simplicity. Generalization to other domains with a distance function (for example strings with a similarity measure, or a semantic distance) can be considered. The data owner distinguishes those numerical values that can be modified for watermarking, denoted in the sequel by weights. Our watermarking protocols will modify weights of a database instance, while leaving other values unchanged. We model such database instances by weighted structures [Grädel and Gurevich 1998]. Let $\tau$ be a signature (or database schema), that is a finite set of relation symbols $\{R_1, \ldots, R_t\}$, with respective arity $r_1, \ldots, r_t$. A finite structure $\mathcal{G} = \langle \mathcal{U}, R_1, \ldots, R_t \rangle$ (or database instance) is an interpretation of each relation symbol of the schema $\tau$ on a finite universe $\mathcal{U}$. We denote by $\text{STRUCT}[\tau]$ the set of all $\tau$-structures. For a given $s \in \mathbb{N}$, a weighted structure $(\mathcal{G}, W)$ is defined by a
finite structure $\mathcal{G}$ and a weight function $\mathcal{W}$, which is a partial function from $U^s$ to $\mathbb{N}$, that maps a $s$-tuple $b$ to its weight $\mathcal{W}(b)$.

**Example 2.1.** The owner of the previous instance (Example 1.1) chooses a unique weight attribute “price”. The corresponding structure is $\mathcal{G} = \langle U, \text{Route}, \text{PriceTable} \rangle$, with, as an example of possible tuples $(\text{TourNepal}, T4) \in \text{Route}, (T1, \text{Paris}, \text{Delhi}, \text{plane}) \in \text{PriceTable}$ and $\mathcal{W}(T1) = 35$.

**Measure of distortion.** Our watermarking algorithms will introduce perturbations into the structure’s weight function $\mathcal{W}$, and these perturbations must be under control of the data owner, so that the data set quality is preserved. The finite part of the weighted structure will remain unchanged: hence watermarking is equivalent to mapping the weighted structure $(\mathcal{G}, \mathcal{W})$ to a new structure $(\mathcal{G}, \mathcal{W}')$.

First, we limit the alteration of weights. Given a constant $c \in \mathbb{N}$, a weighted structure $(\mathcal{G}, \mathcal{W}')$ is said to be a $c$-local distortion of $(\mathcal{G}, \mathcal{W})$ if and only if for all $\bar{b} \in U^s$, $|\mathcal{W}'(\bar{b}) - \mathcal{W}(\bar{b})| \leq c$.

Second, we take into account the impact of distortions on the result of queries. As a query language we will consider for example first-order (FO) or monadic second-order (MSO) formulas. First-order formulas are built from atomic formulas on the database schema with equality, and are closed under classical boolean connectives $\land, \lor, \neg$ and quantifiers $\exists, \forall$. In monadic second-order logic, quantification is also on sets of elements.

Since queries are performed by data servers upon final users’ inputs, we are interested in parametric queries. A formula with parameter $\bar{a}$ is a formula $\psi(\bar{a}, \bar{v})$ with two distinguished variable vectors, $\bar{u}$ and $\bar{v}$, with arity $r$ and $s$ respectively. Given a structure $\mathcal{G}$, let $\psi(\bar{a}, \mathcal{G}) = \{ \bar{b} \in U^s | \mathcal{G} \models \psi(\bar{a}, \bar{b}) \}$. Variables $\bar{a}$ can be assigned to a value $\bar{a}$ by a final user wishing to obtain the result of $\psi(\bar{a}, \mathcal{G})$.

Let $\psi(\bar{a}, \bar{v})$ be the parametric query to be preserved (we focus on the preservation of a unique query $\psi$, but extension to several queries $\psi_1, \ldots, \psi_k$ is straightforward by projection techniques). The weight of a tuple extends to the weight of a query: given a weighted structure $(\mathcal{G}, \mathcal{W})$ and a parametric query $\psi(\bar{a}, \bar{v})$, its weight $\mathcal{W}(\psi(\bar{a}, \mathcal{G}))$ for parameter $\bar{a}$ is the sum of weights of its tuples:

$$\mathcal{W}(\psi(\bar{a}, \mathcal{G})) = \sum_{\bar{b} \in \psi(\bar{a}, \mathcal{G})} \mathcal{W}(\bar{b}).$$

**Example 2.2.** For the database instance in Example 1.1, we consider the first-order query

$$\psi(u, v) \equiv \exists x_1 x_2 x_3 \text{Route}(u, v) \land \text{PriceTable}(v, x_1, x_2, x_3),$$

extracting, given a travel $u$, the set of transports $v$ part of this travel. Summing their corresponding weights models the total cost of travel $u$. We obtain the following
weights:
\[
W(\psi(\text{India discovery}, G)) = W(T1) + W(T2) = 35 + 20 = 55,
\]
\[
W(\psi(\text{Nepal Trek}, G)) = 35 + 15 + 30 = 80,
\]
\[
W(\psi(\text{TourNepal}, G)) = 30 + 50 = 80.
\]

These weights correspond to answers of the parametric aggregate query of Example 1.1.

Query preservation can now be defined. Let \(d \in \mathbb{N}\). A weighted structure \((G, W')\) is a \(d\)-global distortion of a structure \((G, W)\) with respect to \(\psi\) if and only if, for all \(\bar{a} \in U'\),

\[
|W'(\psi(\bar{a}, G)) - W(\psi(\bar{a}, G))| \leq d.
\]

Example 2.3. We consider the original instance given in Example 1.1 and the same query \(\psi\). Let \(\text{PriceTable}'\) and \(\text{PriceTable}''\) be two possible distortions of \(\text{PriceTable}\):

<table>
<thead>
<tr>
<th>Route:</th>
<th>transport</th>
</tr>
</thead>
<tbody>
<tr>
<td>India discovery</td>
<td>T1</td>
</tr>
<tr>
<td>India discovery</td>
<td>T2</td>
</tr>
<tr>
<td>Nepal Trek</td>
<td>T1</td>
</tr>
<tr>
<td>Nepal Trek</td>
<td>T3</td>
</tr>
<tr>
<td>Nepal Trek</td>
<td>T4</td>
</tr>
<tr>
<td>TourNepal</td>
<td>T4</td>
</tr>
<tr>
<td>TourNepal</td>
<td>T5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PriceTable:</th>
<th>transport</th>
<th>departure</th>
<th>arrival</th>
<th>type</th>
<th>price</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>Paris</td>
<td>Delhi</td>
<td>plane</td>
<td>35</td>
<td>45</td>
</tr>
<tr>
<td>T2</td>
<td>Delhi</td>
<td>Nawal.</td>
<td>bus</td>
<td>20</td>
<td>30</td>
</tr>
<tr>
<td>T3</td>
<td>Delhi</td>
<td>Kathm.</td>
<td>plane</td>
<td>15</td>
<td>25</td>
</tr>
<tr>
<td>T4</td>
<td>Kathm.</td>
<td>Simikot</td>
<td>plane</td>
<td>30</td>
<td>20</td>
</tr>
<tr>
<td>T5</td>
<td>Kathm.</td>
<td>Daman</td>
<td>jeep</td>
<td>50</td>
<td>10</td>
</tr>
<tr>
<td>T6</td>
<td>Kathm.</td>
<td>Paris</td>
<td>plane</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

\(\text{PriceTable}'\) is a \(c\)-local distortion of \(\text{PriceTable}\) for constant \(c = 10\), but not a \(d\)-global distortion with respect to \(\psi\) for \(d = 10\), because

\[
W(\text{India discovery, PriceTable}') = 75,
\]

and

\[
W(\text{India discovery, PriceTable}) = 55.
\]

\(\text{PriceTable}''\) respects both local and global distortions for \(c = 10\) and \(d = 10\).

2.2 Watermarking protocols and the non-adversarial view model

The goal of query-preserving watermarking is, given an original structure from a class of structures and a query \(\psi\), to produce altered structures that respect the prescribed local and global distortions.
Definition 2.4. A watermarking problem is a tuple \((\mathcal{K}, \psi)\), where \(\mathcal{K}\) is a class of weighted structures and \(\psi\) is a parametric query.

We generalize definitions of Khanna and Zane [2000] for watermarking structured data. A watermarking protocol is a pair of algorithms \(M\) and \(D\), standing for the “marker” and the “detector”, respectively. Marker \(M\) takes as input an original structure and a binary message \(m\) to be hidden in the data, and computes the watermarked version of the original structure. This structure must preserve the result of a prescribed query \(\psi\). Detector \(D\), given a suspect structure \(\mathcal{G}^*\) as input, extracts the hidden message \(m\). In the first part of this paper, we focus on the non-adversarial model, where the suspect server does not alter its data set in order to evade detection (the full, adversarial model will be considered in Section 6). However, even in the non-adversarial model, constructing correct structures preserving \(\psi\) is a combinatorial problem on its own.

Definition 2.5. Given a watermarking problem \((\mathcal{K}, \psi)\), and \(l, d \in \mathbb{N}\), a \((l,d)\)-watermarking protocol preserving \(\psi\) in the non-adversarial model is a pair of algorithms \(M\) and \(D\) such that:

1. \(M\) takes as input an original structure \((\mathcal{G}, W) \in \mathcal{K}\) and a boolean message \(m \in \{0, 1\}^l\) and outputs a 1-local distortion \(\mathcal{G}_m = (\mathcal{G}, W_m) \in \mathcal{K}\) such that \(\mathcal{G}_m\) is a \(d\)-global distortion of \((\mathcal{G}, W)\).

2. Algorithm \(D\), given as input the original structure \((\mathcal{G}, W)\) and the unchanged suspect structure \(\mathcal{G}^* = \mathcal{G}_m\), outputs the hidden message \(m\).

Parameter \(l\) stands for the number of bits to be hidden. Value \(d\) is the maximum acceptable global distortion on structures produced by the marker. Note that the suspect structure \(\mathcal{G}^*\) exactly \(\mathcal{G}_m\), as we consider first the non-adversarial model. Note also that we restrict our attention to 1-local distortions: weights will only be incremented (+1) or decremented (-1). In fact, we will focus on the asymptotic behavior of the number of distinct watermarked structures. There is at most \((2c+1)^{|U^*|}\) different \(c\)-local distortions of \((\mathcal{G}, W)\), and restricting to \(c = 1\) does not alter this number drastically.

The definition is still not satisfactory. Indeed, the above-defined detector has full access to the suspect structure \(\mathcal{G}^*\) during detection. We call this easier case the instance model (considered in Section 7). In a more realistic scenario, the data owner can only access suspect data by acting as a simple data user. By providing parameters \(\bar{a}\) to a suspect server, the owner will obtain answers to the query \(\psi(\bar{a}, \mathcal{G}^*)\). We call this model the view model. In the sequel we denote by \(A(\psi(\bar{a}, \mathcal{G}))\) the set of answers to query \(\psi\) on parameter \(\bar{a}\), along with their weights:

\[
A(\psi(\bar{a}, \mathcal{G})) = \{(\bar{b}, W(\bar{b}))|\bar{b} \in \psi(\bar{a}, \mathcal{G})\}.
\]

Example 2.6. Recalling Example 2.3, we have

\[
A(\psi(\text{India discovery}, \text{PriceTable})) = \{(T1, 45), (T2, 30)\}.
\]

The set of all possible answers is defined by \(A(\psi, \mathcal{G}) = \{\bar{a}, A(\psi(\bar{a}, \mathcal{G}))|\bar{a} \in U^*\}\). This yields the complete definition:

Definition 2.7. Given a formula \(\psi\), and \(l, d \in \mathbb{N}\), a \((l,d)\)-watermarking protocol preserving \(\psi\) in the non-adversarial view model is a \((l,d)\)-watermarking protocol.
preserving $\psi$ in the non-adversarial model such that the detector uses only $A(\psi, G^*)$ as input.

Definition 2.8. A watermarking problem $(K, \psi)$ is said to have a $(l, d)$-watermarking protocol if there exists $l, d \in \mathbb{N}$ and a pair $(M, D)$ that is a $(l, d)$-watermarking protocols preserving $\psi$ for structures in $K$.

2.3 Scalable watermarking protocols

Capacity and active tuples. Of course, one is interested in protocols with a large hiding capacity. The largest message size $l$ is clearly bounded by the size of the structure. But since the suspect data set is only available through queries to a server, watermarks should be hidden into weights of active tuples of the data set, that is in those tuples that are part of a query answer. Otherwise the detector has no way to recover their weights and to extract the hidden message.

Let $W^{G, \psi}$ be the active tuples of $(G, W)$ with respect to $\psi$:

$$W^{G, \psi} = \bigcup_{\bar{a} \in \mathcal{U}} \psi(\bar{a}, G).$$

We will often use notation $W$ for short.

Example 2.9. Recalling Example 1.1, active tuples are $W = \{T1, T2, T3, T4, T5\}$, and $T6$ is inactive.

In the rest of this paper, we will only distort weights of active tuples. As a consequence, there will be at most $|W|$ useful weights to modify, and the maximum watermarking capacity is $|W|$. It is noteworthy that $\psi(\bar{a}, G)$ and $W$ do not depend on the weight function $W$: we can perturb $W$ without modifying $\psi(\bar{a}, G)$ or $W$. Indeed, function $W$ is not part of the vocabulary of $\psi$.

Scalable watermarking protocols. A watermarking problem may have a watermarking protocol for a constant value of $l$. The interesting situation is when $l$ is an increasing function of $|W|$, i.e. the number of hidden bits grows with the number of active tuples of the problem. The best situation would be to hide $|W|$ bits of data, without distorting results of queries at all.

Definition 2.10. A watermarking problem $(K, \psi)$ possesses a scalable watermarking protocol if there exists $0 < q \leq 1$ and a constant $d \in \mathbb{N}$ such that the same pair of algorithms $(M, D)$ is a $(\Omega(|W|^q), d)$-watermarking protocol for $(K, \psi)$.

3. QUERY-PRESERVING WATERMARKING: GENERAL CASE

3.1 Computing the watermarking capacity

Let $(K, \psi)$ be a watermarking problem, and $d \in \mathbb{N}$. Given a weighted structure in $K$, we consider the problem of counting the exact number of $d$-global distortions with respect to $\psi$. We call this problem $\#\text{MARK}(\leq d)$. We do not know the exact time complexity of $\#\text{MARK}(\leq d)$ on the class of all structures, but we show that a simple variation of this problem is as hard as computing the number of accepting paths of any NP Turing machine; hence is complete for the classical complexity class $\#P$ [Valiant 1979]. In the following, we restrict our attention to 1-local distortions that entail a +1 alteration on an initial segment of the active tuples of size $k$. For a
given watermarking problem \((\mathcal{K}, \psi)\) and \(d \in \mathbb{N}\), we call \(#k\text{MARK}(\leq d)\) the number of such structures with global distortion at most \(d\):

**Definition 3.1** \(#k\text{MARK}(\leq d)\) problem.

—INPUT: a weighted structure \((G, W)\) in \(\mathcal{K}\).

—OUTPUT: the number of 1-local, \(d\)-global distortions \((G, W')\) of \((G, W)\) such that, for \(W = \{b_1, \ldots, b_k, \ldots\}\):

\[
\forall i \in \{1, \ldots, k\}, W'(b_i) = W(b_i) + 1.
\]

We prove the following result:

**Theorem 3.2.** For every \(d \in \mathbb{N}\), \((d + 1)\text{MARK}(\leq d)\) is \#P-complete on the class of all structures.

We start by proving the NP-completeness of a class of related problems, that we call Bipartite-Mark\((d)\), for \(d \in \mathbb{N}\). Let \(G = (V_1, V_2, E)\) be a bipartite graph. A mark \(\delta\) for \(G\) is a map from \(V_2\) to \(-1, 0, 1\). For a given mark \(\delta\), the distortion \(\Delta(v)\) of a vertex \(v \in V_1\) is the sum of alterations of its neighbors, that is:

\[
\Delta(v) = \sum_{v' \in V_2 : E(v, v')} \delta(v').
\]

**Definition 3.3** Bipartite-Mark\((d)\) problem.

—INPUT: a bipartite graph \(G = (V_1, V_2, E)\) where \(V_2 = \{v_{21}, v_{22}, \ldots, v_{2(d+1)}, \ldots\}\).

—OUTPUT: true if there exists a mark \(\delta\) such that:

— for all \(i \in \{1, \ldots, d + 1\}\), \(\delta(v_{2i}) = +1\);

— for all \(v \in V_1\), \(|\Delta(v)| \leq d\).

**Lemma 3.4.** For every \(d \in \mathbb{N}\), Bipartite-Mark\((d)\) is NP-complete.

**Proof.** For a given \(d \geq 1\), we proceed by reduction from 3SAT. Let \(x_1, \ldots, x_n\) be a set of boolean variables and \(\{C_1, \ldots, C_N\}\) be a set of clauses with exactly three literals. We will encode this 3SAT instance into a bipartite graph \(G = (V_1, V_2, E)\).

The set \(V_1\) is composed of:

— vertices \(v_{11}, v_{12}, \ldots, v_{1(d+1)}\),

— vertices \(v_{1x}, v_{1x'}, v_{1x''}, v_{1x'''}\), for each variable \(x_i\) in the 3SAT instance,

— vertex \(v_{1C}\), for each clause \(C_i\) in the 3SAT instance.

The set \(V_2\) is composed of:

— vertices \(v_{21}, v_{22}, \ldots, v_{2(d+1)}\) for its first \(d + 1\) elements,

— vertices \(v_{2neg1}, v_{2neg2}, \ldots, v_{2negd+1}\),

— for each variable \(x_i\) in the 3SAT instance, vertices \(v_{2x}\), and \(v_{2x'}\).

We now turn to the description of edges in \(E\). Figure 1 shows the set of gadgets that we are going to use. First, recall that, by definition, any solution to Bipartite-Mark\((d)\) will put a +1 alteration on vertices \(v_{21}, v_{22}, \ldots, v_{2(d+1)}\). We obtain a set of useful -1 alterations by considering the Minus gadget. In order to guarantee distortion at most \(d\) on vertex \(v_{11}\), that is, \(|\Delta(v_{11})| \leq d\), the only possible marks
\(\delta\) have to set a -1 alteration on vertex \(v_{2neg_i}\). We copy this Minus gadget for vertices \((v_1, v_{2neg_1}), \ldots, (v_{(d+1)}, v_{2neg_{d+1}})\) in order to enforce \((d+1)\) alterations -1 on vertices \(v_{2neg_1}, \ldots, v_{2neg_{d+1}}\).

Second, we encode each boolean variable \(x_i\) with Variable gadgets. The first two gadgets above ensures that vertices \(v_{2x_i}\) and \(v_{2\bar{x}_i}\) can only support a 0 or +1 alteration (-1 is not possible because, due to \(v_{2neg_i}\), the distortion would be smaller than -d). The last two Variable gadgets below ensures that the sum of alterations on \(v_{2x_i}\) and \(v_{2\bar{x}_i}\), that is \(\delta(v_{2x_i}) + \delta(v_{2\bar{x}_i})\), is equal to 1 (if \(d = 1\), edges from \(v_{1x''}\) to \(v_2\) are omitted). As a conclusion, any valid mark has to choose between \(v_{2x}\) or \(v_{2\bar{x}}\) for a +1 alteration, and is forced to set a zero alteration on the remaining vertex. This mimics the boolean valuation of variable \(x\).

It remains to encode clauses. The Clause gadget of Figure 1 encodes for example clause \(x_1 \lor \bar{x}_2 \lor x_3\). Each literal \(x\) or \(\bar{x}\) is encoded by an edge to vertex \(v_{2x}\) or \(v_{2\bar{x}}\). Because of the edges to \(v_{2neg_1}, \ldots, v_{2neg_{d+1}}\), at least one vertex encoding literals has to bear a +1 alteration so that the global distortion is greater than \(-d\).

This construction guarantees that a Bipartite-Mark\((d)\) solution corresponds to a \(\#(d + 1)\)Mark\((\leq d)\) solution, and reciprocally. For a solution to the 3SAT instance, it is sufficient to choose \(\delta(v_{2x}) = +1\) for each variables positively valued, and \(\delta(v_{2\bar{x}}) = 0\) otherwise. All other alterations are determined by this initial choice and the graph constraints. Conversely, given a solution of the Bipartite-Mark\((d)\) problem, a satisfying valuation for the 3SAT instance is obtained by valuating positively all variables \(x\) such that \(\delta(v_{2x}) = +1\), and by valuating negatively the remaining variables.

Finally, the encoding of \(N\) clauses on \(n\) variables requires \(d^2 + 4d + 6n + N + 3\) vertices and \(d^2 + 3d + n(4d + 6) + N(d + 4) + 2\) edges. The overall size of the produced graph is then linear in \(n\) and \(N\), and hence in the size of the 3SAT input instance.

As a conclusion, this reduction from 3SAT shows that searching a valid mark for distortion \(d \geq 1\) is an NP-hard problem. It is also in NP by guessing a mark and
checking conditions. Hence the problem is NP-complete. This result generalizes to the special case $d = 0$ with a similar construct.

We can now prove the theorem.

**Proof of Theorem 3.2.** By reduction from Bipartite-Mark$(d)$. We see the input bipartite graph $G = (E, V_1, V_2)$ as a structure, and consider the trivial query $\psi(a, v) = E(a, v)$. We choose a weight function $W$ such that $\forall v \in V_2, W(v) = 0$. Observe that $W = V_2$. Deciding if this structure has a watermarked weight function $W'$ with $d + 1$ alterations $+1$ on its first elements with global distortion less or equal to $d$ is equivalent to the decision of Bipartite-Mark$(d)$. This latter problem is NP-complete, by Lemma 3.4. Furthermore, the used reduction is parsimonious, that is, preserves the number of solutions. Hence, for any $d \in \mathbb{N}$, #($d + 1$)Mark$(\leq d)$ is $\#P$-complete, since counting the number of solutions of a NP-complete problem through parsimonious reductions is $\#P$-complete.

### 3.2 Impossibility results

The difficulty of finding distortions that preserve the weight of all sets $\psi(\bar{a}, \mathcal{G})$ is related to the intricacy of these sets. A powerful tool to study this intricacy is the Vapnik-Chervonenkis dimension. Let $\mathcal{C}$ be a family of subsets of $V$. A set $U \subseteq V$ is shattered by $\mathcal{C}$ if $\mathcal{C} \cap U = 2^U$, where $\mathcal{C} \cap U = \{C \cap U | C \in \mathcal{C}\}$. The VC-dimension $\text{VC}(\mathcal{C})$ of $\mathcal{C}$ with respect to $V$ is the maximum of the sizes of the shattered subsets of $V$, or $\infty$ if the maximum does not exist. For a formula $\psi(\bar{a}, \mathcal{G})$ and a structure $\mathcal{G}$, let $\mathcal{C}(\psi, \mathcal{G}) = \{\psi(\bar{a}, \mathcal{G})|\bar{a} \in \mathcal{G}^r\}$, and $\text{VC}(\psi, \mathcal{G}) = \text{VC}(\mathcal{C}(\psi, \mathcal{G}))$.

We say that $\psi$ has bounded VC-dimension on a class of structures $K$ if there exists $k \in \mathbb{N}$ such that, for all $\mathcal{G} \in K$, $\text{VC}(\psi, \mathcal{G}) \leq k$. We say that the VC-dimension is maximal if, for all $\mathcal{G} \in K$, $\text{VC}(\psi, \mathcal{G}) = |\text{W}^\mathcal{G}, \psi|$, hence when the full set of active tuples is shattered. In this situation, guaranteed watermarking for arbitrary structures, preserving even trivial queries is impossible.

**Theorem 3.5.** In the non-adversarial view model, a problem $(\mathcal{K}, \psi)$ with maximal VC-dimension does not possess a scalable watermarking protocol.

**Proof.** Let $(\mathcal{K}, \psi)$ be a watermarking problem with maximal VC-dimension. With at most $k$ distorted weights, one can produce at most $\sum_{i=1}^{k} \binom{|W|}{i} 2^i \leq (2|W|)^k$ different weighted structures, encoding at most $O(k \ln |W|)$ bits. Suppose that there exists a scalable watermarking protocol encoding $|W|^q$ bits for a given $q$ with distortion $d_0$. Then it must use a mark $\delta$ with at least $h(|W|) = \frac{1}{q} \ln |W|^q$ distortions with the same sign, say $+1$. The function $h$ is increasing with respect to $|W|$ (since $|W|^q > \ln |W|$), and there exists $n_0$ such that $h(n_0) > d_0$.

We now consider a structure $\mathcal{G}_{n_0}$ which universe has $n_0$ elements. The watermarking protocol must add distortion $+1$ to weights from a set of elements $\mathcal{P}$ with $|\mathcal{P}| = h(n_0) > d_0$. Since, by hypothesis, $\text{VC}(\psi, \mathcal{G}_{n_0}) = |\text{W}^\mathcal{G}_{n_0}, \psi|$, there exists a subset $S$ of $\mathcal{U}^*$ of size $|\text{W}^\mathcal{G}_{n_0}, \psi|$ which is shattered by sets in $\mathcal{C}(\psi, \mathcal{G}_{n_0})$. But since sets in $\mathcal{C}(\psi, \mathcal{G}_{n_0})$ are all subsets of $W$, there exists only one possible $S$: $S = W$ (sets in $\mathcal{C}(\psi, \mathcal{G}_{n_0})$ can not shatter sets outside $W$). So the set $W$ is shattered by results of queries, and there exists a tuple $\bar{a}$ such that $\mathcal{P} = \psi(\bar{a}, \mathcal{G})$. Hence distortion on $\psi(\bar{a}, \mathcal{G}_{n_0})$ is greater than $d_0$, which contradicts the hypothesis to have a scalable watermarking protocol.
It is worth noting that this impossibility argument can be followed with even trivial queries:

**Theorem 3.6.** In the non-adversarial view model, there exists an FO formula \( \psi \) and a class of structures that do not possess a scalable watermarking protocol preserving \( \psi \).

**Proof.** It is sufficient to consider the formula \( \psi(u, v) \equiv E(u, v) \) and the class of structures \( \mathcal{G}_n \) with \( 2^n + n \) vertices, and the simple binary relation \( E \) that links the \( i \)th vertex of the first \( 2^n \) vertices to the \( i \)th subset \( W_i \) of the \( n \) last vertices. The set of active tuples is clearly shattered by \( \psi \) and Theorem 3.5 applies.

**Remark 3.7.** Unbounded VC-dimension is not sufficient: one can construct a class of structures \( \mathcal{G}_n \) of size \( n \) where only half of the active tuples are shattered (\( VC(\psi, \mathcal{G}_n) = |W|/2 \)), with a \((|W|/4, 0)\)-watermarking protocol. Consider the class of structures \( \mathcal{G}_n \) with \( 2^{n/2} + 1 + n \) vertices, and the simple binary relation \( E \) that links the \( i \)th vertex of the first \( 2^{n/2} \) vertices to the \( i \)th subset of the \( n/2 \) last vertices, and the \( 2^{n/2} + 1 \)th vertex \( a \) to all of the \( n \) last vertices. The watermarking problem defined by the query \( \psi(u, v) = E(u, v) \) has \( n \) active tuples and unbounded VC-dimension. The last \( n/2 \) vertices of the active tuples are involved only for query \( E(a, \mathcal{G}_n) \). Putting balanced distortions \((+1, -1)\) or \((-1, +1)\) only on their \( n/4 \) weights gives a watermarking protocol encoding \( n/4 \) bits with distortion 0.

### 3.3 Lower bound on distortion

The previous results shows that, as soon that all subsets of size \( d \) of \( W \) can be captured by \( \psi \), there is no scalable watermarking protocol with error smaller than \( d/2 \). This yields the following lower bound:

**Lemma 3.8.** For any \( d \in \mathbb{N} \), for any class of structures \( \mathcal{K} \) and any logic with equality, there exists a query \( \psi \) such that \((\mathcal{K}, \psi)\) does not possess a scalable watermarking protocol with error smaller than \( d \).

**Proof.** For \( d = 1 \), consider the following query:

\[ \psi(u_1, u_2, v) \equiv (v = u_1) \lor (v = u_2). \]

On any class of structures, this query captures all subsets of active tuples of size 2. If a scalable watermarking protocol exists, for large enough structures it must use at least 2 weight alterations of the same sign, say \((+1)\). Hence there exists parameters \((a_1, a_2)\) capturing these two tuples, and distortions exceeds \( d = 1 \), which contradicts the hypothesis. This argument generalizes for any value of \( d \).

### 4. Watermarking While Preserving Local Queries

#### 4.1 Locality of queries

The previous sections show that, on generic structures, even trivial queries do not possess a scalable watermarking protocol. In this section we limit the relationship between tuples in the structure, and use the locality of queries to provide a scalable watermarking protocol. Given a structure \( \mathcal{G} = \langle U, R_1, \ldots, R_t \rangle \), its Gaifman graph [Gaifman 1982] is the new structure \( \langle U, E \rangle \), where \((a, b) \in E\) if and only if there is a relation \( R_i \) in \( \mathcal{G} \) and a tuple \( \tilde{c} \) in \( R_i \) such that \( a \) and \( b \) appear in \( \tilde{c} \).
The distance $d(a, b)$ between two elements $a$ and $b$ is the length of a shortest path between $a$ and $b$ in the Gaifman graph of $G$. If no such path exists, $d(a, b) = \infty$.

Given $a \in U$, $\rho \in N$, the $\rho$-sphere $S_\rho(a)$ is the set $\{b \mid d(a, b) \leq \rho\}$, and for a tuple $\bar{a}$, $S_\rho(\bar{a}) = \cup_{a \in \bar{a}} S_\rho(a)$. Given a tuple $\bar{c} = (c_1, \ldots, c_n)$, its $\rho$-neighborhood $N_\rho(\bar{c})$ is defined as the structure $\langle S_\rho(\bar{c}), R_1 \cap S_\rho(\bar{c})^{r_1}, \ldots, R_t \cap S_\rho(\bar{c})^{r_t}, c_1, \ldots, c_n \rangle$, where $\forall i, R_i$ has arity $r_i$. Let $\approx$ denote isomorphism of structures. We consider the equivalence relation $\approx_\rho$ on elements of a structure $G$ where $\bar{a} \approx_\rho \bar{b}$ if and only if $N_\rho(\bar{a}) \approx N_\rho(\bar{b})$.

Finally, let $ntp(\rho, G)$ be the number of equivalence classes of the relation $\approx_\rho$. We introduce the important notion of the locality rank of a query:

**Definition 4.1.** Given a query $\psi(u_1, \ldots, u_r)$, its locality rank is a number $\rho \in N$ such that, for every $G \in \text{STRUCT}_\tau$ and two $r$-ary tuples $\bar{a}_1$ and $\bar{a}_2$ of $G$, $N_\rho(\bar{a}_1) \approx N_\rho(\bar{a}_2)$ implies $G \models \psi(\bar{a}_1) \Leftrightarrow G \models \psi(\bar{a}_2)$. If no such $\rho$ exists, the locality rank of $\psi$ is $\infty$. A query is local if it has a finite locality rank. A language is local if each of its queries is local.

For example, Gaifman’s theorem [Gaifman 1982] states that every first-order (relational calculus) query is local. The locality rank of a formula $\psi$ is basically exponential in the depth of quantifier nesting in $\psi$, but does not depend on the size of $G$.

**Example 4.2.** We consider a graph instance $G = \langle U, R \rangle$ and the query $\psi(u, v) \equiv R(u, v)$ that enumerates all elements $v$ at distance 1 of element $u$. This query has locality rank 1: it is sufficient to look at a neighborhood of radius 1 around $u$ and $v$ to devise if $G \models \psi(u, v)$. Figure 2 shows $G$ and neighborhoods $N_1(a)$ and $N_1(d)$ of elements $a$ and $d$. Observe that there are 3 distinct neighborhoods of radius 1 (up to isomorphism), and that $N_1(a) \approx N_1(b), N_1(d) \approx N_1(e)$ and $N_1(c) \approx N_1(f)$.

### 4.2 Watermarking and locality

In the sequel, we restrict our attention to structures in $\text{STRUCT}_k[\tau]$: structures with Gaifman graph of bounded degree $k$. An example of such structures is the class of graphs of bounded degree. Our aim is now to prove the following result:
THEOREM 4.3. In the non-adversarial view model, there exists a scalable watermarking protocol preserving any local query \( \psi(u_1, \ldots, u_r, v_1, \ldots, v_s) \) on \( \text{STRUCT}_k[\tau] \), with capacity \( \Omega(|W|) \) and constant global distortion \( r \).

PROOF. Let \( \psi \) be a formula of locality rank \( \rho \), and \( (U, E) \) be the Gaifman graph of \( \mathcal{G} \). Let \( \text{type}(\bar{b}) \in \mathbb{N} \) be the isomorphism type of any element \( \bar{b} \in W \) with respect to relation \( \approx_{\rho} \). We obtain a set \( \mathcal{P} \) of watermarking positions by the following algorithm. Let \( t_0 \) be the isomorphism type of an equivalence class of \( \approx_{\rho} \) on \( W \) with maximal size (that is, \( \forall t \neq t_0, |\{\bar{b} | \text{type}(\bar{b}) = t\}| \leq |\{\bar{b} | \text{type}(\bar{b}) = t_0\}| \)). Since there is a maximal number \( K \) of such isomorphism types, depending only on \( k \), then \( |\{\bar{b} | \text{type}(\bar{b}) = t_0\}| \geq |W|/K \). We build the set \( \mathcal{P} \) of watermarking positions by repeatedly choosing an element \( \bar{b} \) in \( \{\bar{b} | \text{type}(\bar{b}) = t_0\} \) and by removing all elements in its \( 2\rho \)-neighborhood. Since a \( 2\rho \)-neighborhood has at most \( sk^{2\rho} \) elements, the obtained set \( \mathcal{P} \) has size at least \( |W|/Ks^{2\rho} \).

Then, we construct marks \( \delta_{\mathcal{P}} \) according to the set \( \mathcal{P} \) by assigning a zero distortion on tuples outside \( \mathcal{P} \), and by assigning \( +1 \) or \( -1 \) distortions on tuples in \( \mathcal{P} \), such that they sum up to 0:

\[
\sum_{\bar{b} \in \mathcal{P}} \delta_{\mathcal{P}}(\bar{b}) = 0, \text{ and } \forall \bar{b} \notin \mathcal{P}, \delta_{\mathcal{P}}(\bar{b}) = 0.
\]

There is \( 2^{\Omega(|W|)} \) possible such marks \( \delta_{\mathcal{P}} \). We now show that for any tuple \( \bar{a} \), distortion \( \Delta(\psi(\bar{a}, \mathcal{G})) = |W'(\psi(\bar{a}, \mathcal{G})) - W(\psi(\bar{a}, \mathcal{G}))| \) is a constant. First, observe that

\[
\psi(\bar{a}, \mathcal{G}) = (\psi(\bar{a}, \mathcal{G}) \cap \mathcal{P}) \uplus (\psi(\bar{a}, \mathcal{G})/\mathcal{P}),
\]

where \( \uplus \) denotes disjoint union. Since \( \delta_{\mathcal{P}}(\bar{b}) = 0 \) whenever \( \bar{b} \notin \mathcal{P} \), we have \( W'(\bar{b}) = W(\bar{b}) \) and

\[
\Delta(\psi(\bar{a}, \mathcal{G})) = \Delta(\psi(\bar{a}, \mathcal{G}) \cap \mathcal{P}).
\]

Second, for \( B = S_\rho(\bar{a})^* \), the set of \( s \)-tuples in the neighborhood of tuple \( \bar{a} \), we have

\[
\Delta(\psi(\bar{a}, \mathcal{G})) \leq \Delta(\psi(\bar{a}, \mathcal{G}) \cap (\mathcal{P} \cap B)) + \Delta(\psi(\bar{a}, \mathcal{G}) \cap (\mathcal{P}/B)).
\]

Because of the distance between \( \bar{a} \) and elements in \( \mathcal{P} \), the maximal size of \( B \cap \mathcal{P} \) is \( r \) (this occurs when each element \( a_1, \ldots, a_r \) of tuple \( \bar{a} \) is in the neighborhood of an element in \( \mathcal{P} \)).

Notice also that for \( \bar{b}_i, \bar{b}_j \in \mathcal{P}/B \), because of the choice of \( \mathcal{P} \) and because \( \bar{b}_i \) and \( \bar{b}_j \) are not in the neighborhood of \( \bar{a} \), \( (\bar{a}, \bar{b}_i) \) and \( (\bar{a}, \bar{b}_j) \) are isomorphic. Hence, due to the locality of \( \psi \), \( \mathcal{G} \models \psi(\bar{a}, \bar{b}_i) \leftrightarrow \mathcal{G} \models \psi(\bar{a}, \bar{b}_j) \). Then, by the definition of the watermarking method,

\[
\sum_{\bar{b} \in \mathcal{P}} \delta_{\mathcal{P}}(\bar{b}) = 0.
\]

Then

\[
\sum_{\bar{b} \in \mathcal{P}} \Delta(\psi(\bar{a}, \bar{b})) = 0,
\]
and finally

$$\sum_{b \in \mathcal{P}/B} \Delta(\psi(\bar{a}, b)) \leq |B| \leq r.$$ 

The overall distortion $\Delta(\psi(\bar{a}, \mathcal{G}))$ is then smaller than $r$.

The scheme is deployed deterministically as follows. Given $(\mathcal{G}, \mathcal{W})$ and $\psi$, a precomputation phase, done once, generates the set $\mathcal{P}$. Given a word $m$ of length $l$ to hide, the marker returns $(\mathcal{G}, \mathcal{W}_m)$, where $\mathcal{W}_m$ is the $m$th weights distortion of tuples in $\mathcal{P}$.

Once a suspect data server is localized, the detector (acting as a final user), asks the data server for all possible queries, and obtains weights described in $\mathcal{W}^*$. By comparing them with the original ones, the corresponding subset of $\mathcal{W}^*$ is identified and message $m$ is recovered.

The marker performs $O(nkp\rho, |U|)$ isomorphism tests on constant size graphs. The detector checks $O(|W|)$ values by querying the suspect server. Notice that the marker needs only to compute $\mathcal{P}$ once, and can compute from it every watermarked instance. $\blacksquare$

Combined with the lower bound of Lemma 3.8, there exist queries with $r$ parameters with watermarking error at least $r$. Hence the previous result is tight.

5. PRESERVING MSO-QUERIES ON TREES AND TREE-LIKE STRUCTURES

5.1 Trees and automaton-definable queries

In this section we consider the problem of watermarking labeled trees and tree-like structures, while preserving MSO-queries. These structures can easily model XML documents.

**Example 5.1.** Figure 3 shows an XML document with one of its possible 1-local distortion. We consider the following parametric XPATH query:

$$\psi(a, v) = /\text{school/student[name} = a]/\text{exam}$$

that specifies the set of exam results of persons whose first name is $a$. The goal is to watermark the document while preserving the sum of these exam results. This sum is $\mathcal{W}(\psi(\text{Robert}, \mathcal{G})) = 28$ on the original document, and has global distortion 2 on the second.

XML deals actually with unranked trees, but several methods exist to encode them into binary trees (as in [Milo et al. 2003]), so we will restrict our attention to the binary case. We will use Grohe and Turán [2004] notion of definability of a $k$-ary formula by a tree-automaton.

A binary tree is viewed as a $\{S_1, S_2, \preceq\}$-structure, where $S_1, S_2$ and $\preceq$ are binary relation symbols. A tree $\mathcal{T} = (T, S^T_1, S^T_2, \preceq^T)$ has a set of nodes $T$, a left child relation $S^T_1$ and right child relation $S^T_2$. Relation $\preceq^T$ stands for the transitive closure of $S^T_1 \cup S^T_2$, the tree-order relation. Given a subset $U$ of nodes in $T$, we denote by $\text{lea}(U)$ the least common ancestor of all elements in $U$ according to order $\preceq^T$. A weighted tree $(T, \mathcal{W})$ is a tree with a weight assignment $\mathcal{W} : T^* \rightarrow \mathbb{N}$. Given a finite alphabet $\Sigma$, let $\tau(\Sigma) = \{S_1, S_2, \preceq\} \cup \{P_c | c \in \Sigma\}$ where for all $c \in \Sigma$, $P_c$ is a unary symbol. A $\Sigma$-tree is a structure $\mathcal{T} = (T, S^T_1, S^T_2, \preceq^T, (P^T_c)_{c \in \Sigma})$, where
Query-Preserving Watermarking of Relational Databases and XML Documents

![Diagram of XML documents and a possible 1-local distortion](image)

its restriction \( \langle T, S^T_1, S^T_1, \leq^T \rangle \) is an ordered binary tree and for each \( a \in T \) there exists exactly one \( c \in \Sigma \) such that \( a \in P^T_c \). We denote this unique \( c \) by \( \sigma^T(a) \).

We consider trees with a finite number of distinguishable pebbles placed on vertices. For some \( k \geq 1 \), let \( \Sigma^k = \Sigma \times \{0, 1\}^k \). For a \( \Sigma \)-tree \( T \) and a tuple \( \bar{a} = (a_1, \ldots, a_k) \) of vertices of \( T \), let \( T_{\bar{a}} \) be the \( \Sigma_k \)-tree with the same underlying tree as \( T \) and \( \sigma^{T_{\bar{a}}}(b) = (\sigma^T(b), a_1, \ldots, a_k) \), where \( a_i = 1 \) if and only if \( b = a_i \).

A \( \Sigma \)-tree automaton is a tuple \( B = (Q, \delta, F) \). Set \( Q \) is a set of states, and \( F \subseteq Q \) is a set of accepting states. Function \( \delta : ((Q \cup \{\ast\})^2 \times \Sigma) \rightarrow Q \) is the transition function (\( \ast \not\in Q \)). A run \( \rho : T \rightarrow Q \) of \( B \) on a \( \Sigma \)-tree \( T \) is defined as follows. If \( a \) is a leaf then \( \rho(a) = \delta(\ast, \ast, \sigma^T(a)) \). If \( a \) has two children \( b_1 \) and \( b_2 \), then \( \rho(a) = \delta(\rho(b_1), \rho(b_2), \sigma^T(a)) \). If \( a \) has only a left child \( b \) then \( \rho(a) = \delta(\rho(b), \ast, \sigma^T(a)) \) and similarly if \( a \) has only a right child \( b \), \( \rho(a) = \delta(\ast, \rho(b), \sigma^T(a)) \). Finally, a \( \Sigma_{k+s} \)-tree automaton defines a \( s \)-ary query with \( k \) parameters \( B(\bar{a}, T) = \{b \in T^s | B \text{ accepts } T_{\bar{a}}\} \) on each \( \Sigma \)-tree \( T \). Let \( W = \bigcup B(\bar{a}, T) \).

It is well known that MSO-sentences and tree-automata have the same expressive power. For formula with free variables, a \( \Sigma_k \)-tree automaton is equivalent to an MSO-formula \( \psi(u_1, \ldots, u_k) \) of vocabulary \( \tau(\Sigma) \) if for all \( \Sigma \)-tree, \( B(T) = \psi(T) \).

LEMMA 5.2 [GROHE AND TURÁN 2004]. For any MSO-formula \( \psi(u_1, \ldots, u_k) \) of vocabulary \( \tau(\Sigma) \) there exists a \( \Sigma_k \)-tree automaton \( B \) that is equivalent to \( \psi \).

5.2 Preserving MSO-queries

Our final goal is now to prove the following theorem:

**Theorem 5.3.** In the non-adversarial view model, there exists a scalable watermarking protocol preserving any MSO-definable query on trees or classes of structures with bounded clique-width or bounded tree-width.

To prove this result, we first prove the following theorem, in order to apply Lemma 5.2.

**Theorem 5.4.** In the non-adversarial view model, given a query \( \psi \) defined by
a tree automaton with $m$ states, there exists a scalable watermarking protocol preserving $\psi$ with capacity $\frac{|W|}{4m}$.

We begin by the following lemma:

**Lemma 5.5.** Let $\mathcal{B}$ be a $\Sigma_2$-tree automaton with $m$ states. Then for every $\Sigma$-tree $T$, there exists $n = \frac{|W|}{4m}$ distinct sets $V_1, \ldots, V_n \subseteq W$ and $n$ distinct pairs $(b_i, b_i') \in V_i^2$ of distinct elements such that $\forall i \neq j, V_i \cap V_j = \emptyset$, and $\forall a \in T$:

$$a \notin V_i \rightarrow (b_i \in \mathcal{B}(a, T) \iff b_i' \in \mathcal{B}(a, T)).$$

**Proof.** Informally, we will iterate a construct from Grohe and Turán [2004]: from the bottom-up, we form $|W|/4m$ subtrees of $T$ of size at least $2m$. Since the automaton has only $m$ states, one can find in each $V_i$ a pair of vertices such that the automaton ends in the same state on a given subtree, for all $a \notin V_i$. Then, the automaton can not distinguish between these elements. By using a pair marking on such pairs, the overall distortion is controlled.

More formally, from the bottom-up of $T$, let $U_1$ be a minimal subtree with respect to inclusion with at least $2m$ elements. Since $T$ is binary, $U_1$ contains at most $4m$ elements. We can repeat this construct $2n = \lceil |T|/4m \rceil$ times, obtaining sets $U_1, \ldots, U_{2n}$.

We consider the binary relation $F$ on $H = \{U_1, \ldots, U_{2n}\}$ to be the set of all pairs $(U_i, U_j)$ such that $\text{lca}(U_i) \preceq_T \text{lca}(U_j)$, and there is no $k$ such that $\text{lca}(U_i) \preceq_T \text{lca}(U_k) \preceq_T \text{lca}(U_j)$. Then $(H, F)$ is a forest with $2n$ vertices and at most $2n - 1$ edges. Therefore there is at most $n$ elements of this forest with more that 1 child. Without loss of generality, suppose that $U_1, \ldots, U_n$ have at most one child.

If $U_i$ has no child, let $V_i = \{v \in T|\text{lca}(U_i) \preceq_T v\}$, the elements of the subtree of $T$ rooted at $\text{lca}(U_i)$. If $U_i$ has one child $U_j$, then let $V_i = \{v \in T|\text{lca}(U_i) \preceq_T v$ and $\text{lca}(U_j) \preceq_T v\}$ the set of all vertices of the subtree of $T$ rooted at $\text{lca}(U_i)$ that are not in the subtree rooted at $\text{lca}(U_j)$. Observe that $V_1, \ldots, V_n$ are pairwise disjoint.

Let $1 \leq i \leq n$. If $U_i$ has no child, since $|U_i| \geq 2m$ is greater than the number of automaton states, there exists two distinct elements $b_i, b_i' \in U_i$ such that:

For all $a \notin V_i$, automaton $\mathcal{B}$ running on $T_{ab_i}$ or $T_{ab_i'}$ reaches $\text{lca}(U_i)$ in state $q_i$.

Now if $U_i$ has a child $U_j$, and $q_1, \ldots, q_m$ are the states of $\mathcal{B}$, we define pairs $b_{i,k}, b'_{i,k}$ for $1 \leq k \leq m$ by induction on $k$. Suppose $1 \leq k \leq m$ and that $b_{i,l}$ and $b'_{i,l}$ are already defined for $l < k$. Since $|U_i| \geq 2m$ we have $|U_i \setminus \{b_{i,1}, \ldots, b_{i,k-1}\}| > m$. Therefore there exists distinct elements $b_{i,k}, b'_{i,k} \in U_i \setminus \{b_{i,1}, \ldots, b_{i,k-1}\}$ such that:

There is a state $q_{i,k}$ of $\mathcal{B}$ such that if $a \notin V_i$, the automaton running on either $T_{ab_i}$ or $T_{ab_i'}$ and leaving $\text{lca}(U_j)$ in state $q_k$ reaches $\text{lca}(U_i)$ in state $q_{i,k}$.

Finally, if $U_i$ has no child, and $a \notin V_i$, $\mathcal{B}$ accepts $T_{ab_i}$ if and only if $\mathcal{B}$ accepts $T_{ab_i'}$. If $U_i$ has one child $U_j$, then $\mathcal{B}$ ends in $\text{lca}(U_j)$ in state $q_t$, $\mathcal{B}$ accepts $T_{ab_{i,t}}$ if and only if $\mathcal{B}$ accepts $T_{ab'_{i,t}}$. $\square$

We now claim that pairs $(b_i, b_i')$ are good candidates for a watermarking algorithm.

ACM Transactions on Database Systems, Vol. V, No. N, Month 20YY.
Proof of theorem 5.4. Let \((b_1, b'_1), \ldots, (b_n, b'_n)\) be the \(n\) distinct pairs of Lemma 5.5, \(W'\) be a subset of \(\{1, \ldots, n\}\), and consider a mark that applies \((+1, -1)\) or \((-1, +1)\) distortions on pairs \((b_i, b'_i)\). There are \(2^n\) such marks that sums all to zero. For a \(\Sigma_2\)-tree automaton, let \(a \in T\). Suppose there is a \(j\) such that \(a \notin V_j\). Notice that in this case \(j\) is unique. Then for all \(i \neq j\), \(a \notin V_i\) and \(B\) accepts \(T_{ab}\), if and only if it accepts \(T_{ab'}\). Since distortion on weights \(b_i\) and \(b'_i\) is zero, distortion on \(W(B(a, T))\) is limited by the pair \(b_j, b'_j\). This distortion is at most 1. Otherwise, if \(\forall i, a \notin V_i\), then the induced distortion of all pairs is 0. Finally, the number of pairs is \(|W|/4m\). This result generalizes to a \(\Sigma_{r+s}\)-tree automaton with distortion at most \(r\).

The time complexity of this method corresponds to a full traversal of the tree \(T\) to form subsets of size at most \(4m\), and for each set, to a search of pairs of nodes that force the automaton to end in the same state. It takes at most \(|T|\) steps to visit the tree and \((4m)^2\) runs of the automaton on \(|T|/4m\) sets. The overall complexity is then \(O(m|T|)\).

We can now end with the proof of the main theorem.

Proof of theorem 5.3. We prove the existence of a scalable watermarking protocol preserving MSO-formulas on (a) tree structures, (b) bounded clique-width structures and (c) bounded tree-width structures:

(a) Let \(\psi\) be an MSO-formula on trees. By applying Lemma 5.2, we obtain an automaton \(B\) equivalent to \(\psi\). Then, by Theorem 5.4, there is a corresponding scalable watermarking protocol preserving \(B\), hence \(\psi\), on trees.

(b) For structures with bounded clique-width, we apply Lemma 16 of [Grohe and Turán 2004]: given a structure \(G\) with bounded clique-width, we can construct a labeled parse-tree \(T\) such that, for any MSO-formula \(\psi(\bar{a})\) there exists a MSO-formula \(\tilde{\psi}(\bar{a})\) such that \(\psi(G) = \tilde{\psi}(T)\). This reduce the bounded clique-width case to the tree case and, by step (a) of this proof, there exists a scalable watermarking protocol preserving \(\tilde{\psi}\), hence \(\psi\), on bounded clique-width structures.

(c) Finally, structures with bounded tree-width \(k\) have a bounded clique-width of at most \(2^k\). Then, by step (b) of this proof, there exists a scalable watermarking protocol preserving MSO-formulas on bounded tree-width structures.

If we lower the restriction on the class of structures, scalability is lost:

Theorem 5.6. In the non-adversarial view model, there exists an MSO formula \(\psi\) and a class of structures with unbounded tree-width that do not possess a scalable watermarking protocol preserving \(\psi\).

Proof. Example 19 in [Grohe and Turán 2004] exhibits a MSO formula \(\psi\) with unbounded VC-dimension on the class of grids, which has unbounded tree-width. This shows actually that for all grid \(G\), the set \(\bigcup \psi(\bar{a}, G)\) is shattered by sets in \(\{\psi(\bar{a}, G) : \bar{a} \in \bar{a}'\}\). The corresponding watermarking problem with the same formula \(\psi\) on the same class of structures is such that for all \(G\), \(W^{\psi, G} = \bigcup \psi(\bar{a}, G)\).
Hence for all $G$, its set of active tuples is shattered, so $VC(\psi, G) = |W^{G, \psi}|$, showing by Theorem 3.5 that no scalable watermarking protocol is possible. □

6. ADVERSARIAL MODEL

Up to now, we considered a non-adversarial model. In the adversarial model, data servers can perform any reasonable distortion on the watermarked structure $G_m$, or on their query answers. The reasonable limit of the global distortion is denoted $d'$. Once data servers are allowed to perform distortions, deterministic detectors are no longer possible: as a matter of fact, an unstoppable strategy for the attacker is to guess the inserted mark and its position, and to modify weights accordingly. Hence we consider detectors with a controlled probability of error $\delta$, where $\delta$ is chosen by the data owner (with for example $\delta = 10^{-6}$).

More formally, a probabilistic algorithm has the ability to pick a random bit $b$ at each step, and to adapt its computation according to the value of $b$. Hence a given computation is a path in the tree of all possible random choices along with its corresponding probability: both form a probability space $\Omega$. We consider a watermarking protocol that may succeed with high probability and may fail (stop and abandon), or produce an incorrect result with a small probability:

**Definition 6.1.** Given $0 \leq \delta < 1$ and $l, d, d' \in \mathbb{N}$, a $(l, d, d')$-watermarking protocol preserving $\psi$ in the adversarial view model is a $(l, d)$-watermarking protocol preserving $\psi$ such that the detector, using only answers $A(\psi, G^*)$ from a $d'$-global distortion $G^*$ of $G_m$, verifies:

$$\Pr_{\Omega}[D \text{ outputs } m] \geq 1 - \delta.$$

Khanna and Zane [2000] proposed a general technique to turn a non-adversarial protocol into an adversarial one. Their result is quite general as it applies to their edge model (that corresponds to our view model), and also to their distance model, that corresponds to a detector using only the aggregate value $W(\psi(\bar{a}, G))$ of the query answer, and not all the set $A(\psi(\bar{a}, G))$ (we will elaborate on this distinction in the conclusion section). We recall their result here for completeness, adapted to our view model, and refer the reader to the original paper for a more precise exposition.

Two natural hypothesis are used to constraint the behavior of the attacker. Let $(G, W)$ be an original structure, $(G, W')$ be a watermarked version, and $(G, W^*)$ be a voluntary alteration of $(G, W')$. The first assumption indicates that there is a limit to the distortion the attacker can add to a structure, imposed by its intended use:

**Assumption 1 Bounded distortion.** The attacker respects the global distortion assumption, for an absolute constant $d'$, that is, for all $\bar{a} \in U^*$,

$$|W^*(\psi(\bar{a}, G)) - W(\psi(\bar{a}, G))| \leq d'.$$

It is noteworthy that there is no relationship between $d'$ and the owner maximal distortion $d$. This allows a large freedom to the attacker.

Let $S$ be the set of marks used by the marker algorithm, and $Q$ be the set of tuples queried by the detector. The second assumption guarantees the probabilistic
independency between the set of queries $Q$ and the set of marks $S$. Given a mark $\delta : W \to \mathbb{Z}$, we define $W_\delta$ such that, for all $b \in W$, $W_\delta(b) = W(b) + \delta(b)$.

**Definition 6.2.** A set $Q \subseteq W$ is low-bias with respect to $S \subseteq \{-1, 0, +1\}^W$ if for all $\bar{a} \in Q$,

$$|\delta(\bar{a})| \leq 1 \quad \text{and} \quad \forall z \in \{-1, 0, +1\}, \Pr[\delta(\bar{a}) = z] \leq \frac{1}{2}.$$ 

**Definition 6.3.** Let $\gamma, p > 0$. A set $S \subseteq \{-1, 0, +1\}^W$ is $(\gamma, p)$-unpredictable if for any $Q \subseteq W$ such that $Q$ is low-bias with respect to $S$ and $|Q| = \omega(1)$, any strategy $W^*$ available to the adversary satisfies

$$\Pr_{\delta \in S} \left[ \sum_{\bar{a} \in Q} |W^*(\bar{a}) = \delta(\bar{a})| > \left(\frac{1}{2} + \gamma\right)|Q| \right] < p.$$ 

**Assumption 2 Limited knowledge.** The attacker has limited knowledge on the mark distribution of the owner, that is, for any $S \subseteq \{-1, 0, 1\}^W$ such that $|S| = \omega(1)$, $S$ is $(\gamma, p)$-unpredictable.

This second assumption indicates that the attacker does not know exactly what information has been introduced into the structure (and does not know the original, non-marked structure). This models also the situation where a server is indeed not malicious, but uses data from another source, similar to the owner’s database (false positive detection).

Khanna and Zane’s framework is the following. Given a watermarking protocol in the non-adversarial model for $\psi(u_1, \ldots, v_1, \ldots, v_s)$, the framework requires a pair of distinct values $v_1, v_2 \in \{-1, 0, 1\}$, a set of marks $S \subseteq \{-1, 0, 1\}^{|W|}$, and a set of query parameters $Q = \{\bar{a}_1, \ldots, \bar{a}_L\} \subseteq \mathcal{U}^s$ such that:

1. $Q$ and $S$ have a significant size, that is $|Q|, |S| \geq \omega(1)$;
2. $Q$ is low bias with respect to $S$;
3. The allowed global distortion is $d$, that is, for all $\delta \in S$, for all $\bar{a} \in \mathcal{U}^s$,

$$|W_\delta(\psi(\bar{a}, G)) - W(\psi(\bar{a}, G))| \leq d.$$ 

4. Any set of query answers can be mapped to a unique $\delta \in S$: for any $D : Q \to \{v_1, v_2\}$, there is a unique $\delta \in S$ such that

$$\forall \bar{a} \in Q, W_\delta(\bar{a}) = W(\bar{a}) + D(\bar{a}).$$

**Theorem 6.4.** [Khanna and Zane 2000] Under the Bounded distortion and Limited knowledge assumptions, given $\gamma < \frac{1}{(d+1)}$ and $p \geq e^{-\alpha(L)}$, any non-adversarial scalable watermarking protocol consistent with the framework can be used to encode $O(L)$ bits in the adversarial model, with a detector’s error probability at most max\{2p, o(1)\}.

Observe that the watermarking robustness is obtained by lack of knowledge (an attacker knows there is a mark, but do not know its amplitude and distribution) and not by using the intractability of a computational problem, like in the cryptographic setting. All the watermarking protocols presented in this paper comply with Khanna and Zane’s framework, so extend to the adversarial setting.
COROLLARY 6.5. There exists a scalable watermarking protocol in the adversarial view model preserving

—local queries on structures with bounded degree Gaifman graph;
—MSO-queries on trees and structures with bounded clique-width or tree-width.

PROOF. By Theorem 4.3 and 5.3, there exists a non-adversarial scalable protocol for the above logical fragments and classes of structures. Let \( d \in \mathbb{N} \) be the allowed global distortion.

For local languages, by the proof of Theorem 4.3, there is a subset \( \mathcal{P} \subseteq \mathcal{W} \) such that any distortions \( \delta \) restricted on \( \mathcal{P} \) summing up to 0 lead to a correct structure with global distortion smaller than \( d \). We split \( \mathcal{P} \) into disjoint pairs \( V_1, \ldots, V_n \). For any pair \( V_i = \{b_{i1}, b_{i2}\} \), the set of marks \( S \) is obtained by choosing uniformly and independently between \( \{\delta(b_{i1}) = +1, \delta(b_{i2}) = -1\} \) or \( \{\delta(b_{i1}) = 0, \delta(b_{i2}) = 0\} \), for each set \( V_i \). Any such choice also leads to an overall distortion on \( \mathcal{P} \) smaller than \( d \). For MSO-queries, Lemma 5.5 also provides such sets \( \mathcal{P} \) and \( V_1, \ldots, V_n \), suited for the same set of marks \( S \). Let \( v_1 = +1 \) and \( v_2 = 0 \). We choose the set of query \( Q = \{b_{i1}, \ldots, b_{in}\} \), that is the first tuple of each pair.

Because the considered protocols are scalable, \( |\mathcal{P}| = 2n \) is \( \Omega(|\mathcal{W}|) \), hence \( |Q| \) and \( |S| \) are \( \omega(1) \), and moreover \( |S| \) is exponential in \( |Q| \). \( Q \) is low-bias with respect to \( S \) because alterations in \( S \) are chosen independently and uniformly. The maximum distortion implied by any \( \delta \) is smaller than the prescribed \( d \). For any valuation \( D : Q \rightarrow \{v_1, v_2\} = \{1, 0\} \) of recovered values from \( Q \) by the detector, we can find the unique corresponding mark \( \delta \):

\[
\forall b_{i1} \in Q, \delta(b_{i1}) = D(b_{i1}), \delta(b_{i2}) = -D(b_{i1}).
\]

This protocol corresponds to the framework. Hence, by Theorem 6.4, \( O(n) \) bits can be encoded such that the error probability of the detector is at most \( \max\{2p, o(1)\} \). Finally, \( n = |\mathcal{W}|/2 = \Omega(|\mathcal{W}|) \). Thus, there exists an adversarial scalable protocol for the considered logical fragments and classes of structures.

We do not consider here the general problem of collusion attacks, where servers combine several watermarked copies of the database to erase the watermark (a full exposition is given in [Lafaye et al. 2008]).

7. INSTANCE MODEL AND NON-PARAMETRIC QUERIES

In this section we consider two easier models that guarantee a huge watermarking bandwidth.

Instance model. In this model, the detector as full access to the suspect instance: the marker can hide information outside the set of active tuples \( W \). Let \( \psi(u_1, \ldots, u_r, v_1, \ldots, v_s) \) be the query to be preserved. The corresponding watermarking bandwidth then depends on the size of \( W \), relatively to the size of the set \( \mathcal{U}^s \) of all possible tuples:

**THEOREM 7.1.** In the instance model, there exists a \((|\mathcal{U}^s| - |W|, 0)\)-watermarking protocol preserving any query \( \psi(u_1, \ldots, u_r, v_1, \ldots, v_s) \).

**PROOF.** Let \( F = \mathcal{U}^s \setminus W \). Hence, any tuple in \( F \) does not participate in any set \( \psi(\bar{a}, \bar{g}) \), thus the corresponding weights have no impact on any \( W(\psi(\bar{a}, \bar{g})) \).
putting a distortion 0 or (+1) on each tuple in \( F \), \(|F| \) bits can be hidden with global distortion 0. The time complexity related to this protocol corresponds to the computation of the set of active tuples \( W \).

If \(|U^s| - |W|\) is small, that is if \( W \) covers a non-negligible part of \( U^s \), the problem reduces to the view model.

Non-parametric queries. Queries considered up to now have parameters fulfilled by data users. In the present setting, users simply obtain the result of a set of predefined queries without parameters \( \psi(v_1, \ldots, v_s) \) (observe that there is no parameter \( \bar{u} \)):

**Theorem 7.2.** There exists a \((|W|, 0)\)-watermarking protocol preserving non-parametric queries \( \psi(v_1, \ldots, v_s) \).

**Proof.** Let \( \bar{w} \in W \) be an active tuple. Being active, it belongs to the image of some \( \psi_i \) on \( G \). Let \( cl(\bar{w}) = \{ i | \bar{w} \in \psi_i(G) \} \) be the class of \( \bar{w} \), that is the set of query indices whose image contains \( \bar{w} \). Observe that two tuples of the same class appear in the same set of queries. Hence, if their alterations sum up to zero, the overall distortion on their common set of queries is also zero. There is at most \( 2^k \) distinct classes. Hence, there is set \( W_0 \subseteq W \) of size at least \(|W|/2^k \) such that \( \forall \bar{w}, \bar{w}' \in W_0, cl(\bar{w}) = cl(\bar{w}') \). Consider any partition of this set into pairs, and any mark that sums up to zero on each pair. This yields \( 2^{|W_0|} \) different marks. The partition size is \(|W|/2^{k+1} \), which is \( \Omega(|W|) \) (this bound can be enhanced to \(|W| - 2^k \) by considering all possible classes, and discarding at most \( 2^k \) tuples because of odd-size classes). Computing the active tuples \( W \) and their respective class can be done in one pass, and the overall time complexity of this protocol is equivalent to the computation of the set \( W \).

8. INCREMENTAL WATERMARKING

In this section we suppose that a data owner needs to update the database and propagates changes to each of the registered data servers. The problem is then to maintain the watermark he has inserted. In the sequel, an update \( u \) is a function that maps a weighted structure \((G, W)\) to an updated structure \( u((G, W)) \). Hence, a sequence of updates \( u_1, \ldots, u_n \) yields a sequence of weighted structures \((G^1, W^1) = u_1((G^0, W^0)), \ldots, (G^n, W^n) = u_n((G^{n-1}, W^{n-1}))\).

**Definition 8.1.** For a class of updates \( U \), a watermarking protocol maintaining \( U \) is a triplet \((\mathcal{M}, \mathcal{M}_U, \mathcal{D})\), where:

- \( \mathcal{M} \) is a marker as in Definition 2.5, producing, given an original structure \((G^0, W^0)\) and an initial word \( m \), a watermarked structure \((G^m_0, W^0_m)\);
- \( \mathcal{M}_U \) is a function that, given an update \( u \in U \) on the owner’s current structure \((G^i, W^i)\), outputs a new watermarked structure \((G^a_m, W^i_m)\);
- \( \mathcal{D} \) is a detector as in Definition 2.5, but for any watermarked structure \((G^i_m, W^i_m)\).

Let \((G, W)\) be a weighted instance. We consider **weights-only updates**: a weights-only update \( u = (\bar{a}, \Delta) \) where \( \bar{a} \in U^s, \Delta \in \mathbb{N} \), is an update that alters only the weight \( \bar{a} \) by the value \( \Delta \), and leaves the finite part unchanged, that is:

- \( u((G, W)) = (G, W') \);
For this class, updating the watermarked instance is easy.

**Theorem 8.2.** Previous watermarking protocols maintain weights-only updates.

**Proof.** First, recall that the set of active tuples, $W$, is not sensitive to the weights values, as its definition rely only on the finite structure $G$. Then tuples involved in queries do not change during an update. Let $\delta$ be the used mark. For a mark distortion

$$W_{i+1}^m(\bar{a}) = W_i^m(\bar{a}) + \delta(\bar{a}),$$

in the current watermarked instance $(G_i^m, W_i^m)$, and a weights-only update $(\bar{a}, \Delta)$ such that

$$W_{i+1}^i(\bar{a}) = W_i^i(\bar{a}) + \Delta,$$

we propagate the same update $\Delta$:

$$W_{i+1}^{i+1}(\bar{a}) = W_i^{i+1}(\bar{a}) + \Delta.$$

The same global distortion is obtained for the new instance. Since the detector extracts the watermark by computing the difference between the watermarked and original weights, for example,

$$W_{i+1}^m(\bar{a}) - W_{i+1}^i(\bar{a}) = (W_i^m(\bar{a}) + \Delta + \delta(\bar{a})) - (W_i^i(\bar{a}) + \Delta) = \delta(\bar{a}),$$

it is only sensitive to the watermark $\delta$, and not the update $\Delta$. Hence, any weights-only update $(\bar{a}, \Delta)$ on the original structure is simply propagated as an update $(\bar{a}, \Delta)$ on the watermarked one. \qed

### 9. PRACTICAL ASPECTS

The proposed protocols are rather efficient: first, the set $W$ of active tuples has to be computed. Then, a set of watermarking positions $P$ has to be computed once, and the proper watermarking and detection operations takes linear time in $|P|$. Finding the set $P$ has a linear time complexity according to the size of the structure $G$: for example, the marker for local languages on structures with bounded-degree Gaifman graph performs $O(ntp(\rho, G)|U_r|)$ isomorphism tests on constant size graphs. The marker for an MSO query $\psi$ on trees requires $O(m|T|)$ steps, where $m$ is the size of the automaton simulating $\psi$. It is known [Stockmeyer and Meyer 1973] that in the worst case, $m$ is in $tower(n)$, where $n$ is the size of the formula $\psi$. The $tower$ function is such that $tower(0) = 1$ and for $i \geq 1$, $tower(i) = 2^{tower(i-1)}$. The main difficulty from the computational point of view is then the role of these huge constants.

Similarly, while error bounds presented in this work elaborate on previous ones, the watermarking rate suffers from the involved constants. Indeed, the watermarking rate is $|W|/Ksk^\rho$, where $K$ is the number of possible isomorphism type of connected graphs with at most $k$ vertices, and $\rho$ is the locality rank of the query, which is exponential in the number of quantifier alternation of $\psi$. But the method of compensating distortions is still fruitful in the practical context. As shown in [Lafaye et al. 2008], a heuristic search of pairs is likely to succeed, yielding only...
a minute distortion. In this latter work, several practical questions not considered in the present work were resolved. We mention them here for completeness:

—Identification of tuples weights, using key dependencies (in the present protocol, the owner has to specify its instance as a finite part and a weighted part, which is not always trivial);

—Blindness of the protocol: the present protocol required the original data set for detection, which is not always affordable. The $(+1, -1)$ pairing was enhanced to an exchange of distinct bits between two values;

—Robustness against collusion attack: malevolent data servers may compare their watermarked data set in order to locate differences. By choosing marks in a collusion-secure fingerprinting code [Tardos 2008], robustness against a prescribed collusion size in guaranteed.

10. CONCLUSION AND FUTURE WORK

In this paper we considered the problem of watermarking databases or XML documents, while preserving a set of queries in a specified language $L$. We gave structural arguments for the existence of a watermarking protocol related to the VC-dimension of sets definable in $L$. We showed that watermarking on arbitrary instances is impossible, and that languages and structures with bounded VC-dimension established by Grohe and Turán have also good watermarking properties. But we do not know if bounded VC-dimension is a sufficient condition to obtain a scalable watermarking protocol.

This work can be extended in various directions. First, one can elaborate on the error model: we considered an absolute error on local and global distortions, while classical studies from the approximation literature consider relative errors rather than absolute ones, because relative approximation is preserved under composition. Observe however that a relative perturbation $1 \pm \varepsilon$ of weights always yields a global distortion of at most $1 + \varepsilon$. Hence the watermarking problem becomes trivial. But a mix between local absolute error and global relative error can be considered.

Second, our model is limited to the instance and view models. The adversarial framework from Khanna and Zane [2000] adapts also to a more general aggregate view model, where only the numerical results of queries is needed to perform detection. We think that, by controlling the impact of alterations on queries for each parameter, protocols for this aggregate model could be obtained for the same logical fragments and classes of structures we considered here.

Third, our model does not capture exactly the result from Khanna and Zane [2000] since shortest path queries are indeed an optimization problem (notice however that the VC-dimension of weighted graphs with respect to their shortest path is bounded). Optimization has received a large interest from the finite model theory community [Kolaitis and Thakur 1994; Papadimitriou and Yannakakis 1991]. An interesting point is to find relationships between logical definability of such problems, mainly their weighted versions [Zimand 1998], and their watermarking capacity.

ACM Transactions on Database Systems, Vol. V, No. N, Month 20YY.
ACKNOWLEDGMENTS

I would like to thank Cristina Bazgan for helpful remarks on the proof of Lemma 3.4, and Caroline Fontaine for her watermarking expertise. I also thank the ACM TODS editors and anonymous reviewers, for their many constructive comments.

REFERENCES


ACM Transactions on Database Systems, Vol. V, No. N, Month 20YY.
Query-Preserving Watermarking of Relational Databases and XML Documents


Received Month Year; revised Month Year; accepted Month Year