Watermarking Maps: Hiding Information in Structured Data

Sanjeev Khanna * Francis Zane†

Abstract

Watermarking has become an important technique for providing copyright protection for images, audio, VLSI designs, and many other types of data. Hidden information is embedded in data which allows the owner of the data to trace pirated copies, identifying a guilty party.

We consider here the problem of watermarking maps. Our work is motivated by the following setting: MapQuest generates a database of map information (represented as a graph) and sells differently modified copies to several web sites like Yahoo!. These web sites respond to end-user queries which ask how far apart two locations are, or how to drive from one to the other. When MapQuest encounters a web site answering such queries, it would like to ascertain if the site is using a copy of its database, and if so which copy, all based only on how the site responds to its queries.

This setting introduces two new and interesting problems for watermarking. First, because of the structure of shortest paths in a graph, modifications are not independent, and even small modifications can accumulate to a large overall distortion. Second, access to the data being tested is indirect, with access only provided via queries.

We study this problem in various abstract settings that highlight the unique aspects of watermarking maps as well as give insights into the underlying combinatorial structure. Building on the insights from simple models, we design and analyze a watermarking scheme in a general setting which corresponds naturally to the example above. In the context of this more realistic model, we show that a map owner can distribute a large number of copies introducing only a negligible distortion, and further that sites using pirated data are likely to be detected even if they attempt to conceal their identity.

1 Introduction

With the widespread availability of digital information, the problem of protecting that information from illicit copying has gained new prominence. A widely used method for addressing this problem is *watermarking* (also called *fingerprinting* [11]), where one embeds hidden information into the data which encodes ownership and copyright information. Data suspected of being pirated can then be tested for this hidden information, determining if the data is copyrighted. Furthermore, by embedding unique hidden information into each copy, pirated data can be traced back to its original purchaser.

This approach has been applied to a wide range of media: images [3, 4], audio [1], video, etc. More recently, intellectual-property protection problems related to VLSI design, such as placement and routing [5] and implementations of finite state machines [7], have also been addressed in this fashion. In addition, there are cryptographic protocols which address related issues, such as how to distribute copies with guarantees of anonymity and reliable authentication [8, 9].

In this paper, we study the application of watermarking techniques to protect map data. Consider the following situation. Through expensive surveys, a company, referred to as the *owner* of the map, compiles accurate map data, represented by a weighted graph where nodes represent locations, edges represent links between locations, and the weights represent the distance between adjacent locations. This data is then sold to other parties called *providers* who provide *end-users* access to it. This access is generally indirect: the provider allows the end-users to query pairs of nodes, and responds to each query with appropriate information related to the pair, such as the distance between the nodes, or a shortest route, or both. This situation already exists for maps of the US roadway system, with companies such as MapQuest acting as map owners, web sites like Yahoo! acting as providers, and end-users across the Internet. In addition, it is not hard to envision similar situations in which this could arise, such as providing routing information on a network.

The Map Watermarking Problem: The goal of this paper is to present *watermarking schemes* (marking schemes) which allow the owner to distribute and identify many different copies of his original graph G, with associated edge lengths $\ell(e)$, in order to protect this map data. Each time the owner sells a copy of the map to a provider, he distributes a copy with a unique set of modifications, refered to as a *marked copy*, which consists of the original graph G with new edge lengths $\ell'(e)$. Gross modifications to the structure of the graph, such as introducing new nodes or edges, are not allowed, since the resulting marked copies are clearly incorrect. At the same time, changes in length which are *small*

*Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA 19104. This work was done when the author was at Bell Laboratories, Murray Hill, NJ.
Email : sanjeev@cis.upenn.edu.

†Department of Fundamental Mathematics Research, Bell Labs, 700 Mountain Avenue, Murray Hill, NJ 07974.
E-mail: francis@research.bell-labs.com.
URL: http://cm.bell-labs.com/cm/ms/who/francis.

compared to the natural unit of length are not allowed, since such changes can be easily undone, even unintentionally, by simple rounding.

Each marked copy must involve only a slight distortion with respect to G and ℓ, measured in terms of the maximum error in inter-point distances, since badly-distorted data is not acceptable to the providers or the end-users. When confronted with a provider suspected of using the data illegally, an owner should be able to accurately determine the unique provider from whom that copy must have originated using only publicly accessible information. That is, the owner accesses the provider's data as an end-user, makes queries, and analyzes the responses.

Here, we consider a number of different models, depending on what distortion is allowed, what information is returned in response to a query, and to what extent the adversary may further distort the data to evade detection. Our goal throughout is to maximize the number of copies the owner can distribute subject to these constraints, and the main difficultly is that these constraints put limits on the number of distinct, distinguishable marked copies.

The requirements that marked copies must have small distortion and that copies must be distinguishable via queries, present new problems quite different from those studied previously. Although the problem of watermarking while introducing small distortion has been studied extensively in the case of media watermarking, the use of a concrete distortion measure like interpoint distance error makes it nontrivial even to identify a large space of small-distortion changes. For example, small local changes, if not made carefully, may accumulate into a large distortion over the whole graph. By contrast, for media documents, such as images, the relevant distortion measure is based on human perception, and can only be quantified in an ad-hoc way. In [6], a model is proposed for analyzing media watermarking schemes, but the abstract distortion measure used (the L_2 difference of the change in some vector representation) oversimplifies the problem of finding small-distortion changes.

The second goal is a new one, largely because the problem does not exist in the well-studied media watermarking case. If the owner encounters a suspect copy of an image, then he has complete access to the image data for this copy, which he can then use to determine the guilty party. Because in our problem the owner must distinguish the copies on the basis of queries alone, this imposes an additional limitation on the owner, forcing him to take this into account when modifying the graph. This issue seems likely to appear in other settings: one might like to determine how a chip was designed or produced given access to the chip but not the proprietary design data, for example.

Related Work: Previous theoretical work on watermarking [2, 6] has largely centered on providing resistance to *collusive attacks*, which compare several differently marked copies an attempt to defeat the scheme. By contrast, our focus here will be on *single copy* attacks, detecting providers who have access only to a single marked copy. Collusive attacks are of course more general. However, the models previously used to study watermarking problems *require* collusive attacks. In the model used by [2], attacks explicitly result from comparing different copies; in [6], the distortion constraint used implies that large distortions can only be undone by averaging several different copies. Here, the map watermarking model is sufficiently rich to make even analyzing single-copy attacks an interesting problem.

Finally, some very recent work on watermarking VLSI designs is aimed at similar structured problem domains. For example, in [10], information is encoded by producing different solutions to 3-coloring instances, where we encode information by producing different shortest paths between points. However, our aims are quite different. In VLSI watermarking, the goal is generally to protect the intellectual value of the optimization software, encoding information by choosing among different near-optimal solutions. Here, the optimization (shortest-path computations) is trivial, and the goal is to protect the data which is being operated on. This problem seems much more amenable to theoretical analysis, since the solution spaces of the NP-hard problems of interest in VLSI seem difficult to characterize.

Outline: We begin by presenting our models and notation in Section 2 and then survey our results in these models in Section 3. Following that, we focus on models capturing our motivating example, and prove results relevant to them. In Section 4, we start by considering a model in which the provider's behavior is restricted; then in Section 5, we demonstrate how to generalize this to other settings.

2 Preliminaries

We now formalize the watermarking problem for maps and describe the various models that we study here. A model is characterized by three parameters:

- How is distortion measured?
- What type of information does the provider give in response to queries?
- Is the provider free to answer as he chooses in order to evade detection?

In what follows, we describe these parameters in detail. We start with some notation.

2.1 Notation

The input data is always assumed to be represented as a *connected* graph $G = (V, E)$ with a *length* function $\ell : E \to \mathcal{Z}$. For a node $u \in V$, we let $N(u)$ denote its neighbors in G. For any pair of nodes $u, v \in V$, $d_G(u, v)$ denotes the shortest distance between u and v, as determined by the metric induced by ℓ. We denote by $\mathcal{P}(G)$ any collection of $\binom{n}{2}$ paths $P_1, P_2, ...$, such that there is a shortest path for each pair of nodes. Given a graph $G = (V, E)$ with length function ℓ, a *marked copy* of G is another graph $G'(V, E)$ with length function ℓ'. Note that since the values $\ell(e)$ and $\ell'(e)$ are assumed to be integral, all modifications to e change its length by at least 1. We use m and n to denote the number of edges and nodes, respectively, in the input graph. Finally, given a path $P = \{e_1, \ldots, e_k\}$ in a graph with length function ℓ, let $\ell(P) = \sum_{i=1}^{k} \ell(e_i)$ be the length of the path.

Remark: If one allows very small changes to the edge lengths, the marking problem becomes easy. However, such changes are obvious (unless the data is extremely high precision) and fragile (destroyed by data conversion and rounding), which makes solutions based on such changes impractical.

2.2 Measures of Distortion

When creating a marked copy of an input graph G, the marking scheme modifies the length function to encode the identity of a provider. Since we would like to ensure that the quality of information provided by a marked copy is close to that of the original, we require that these modifications do not introduce a large distortion in the marked copy G'. There are two natural measures of distortion, additive and multiplicative.

We say that G' has an *additive distortion* of d if $\max_{u,v \in V} |d_G(u, v) - d_{G'}(u, v)| \leq d$. We say that G' has a *multiplicative distortion* of d if $\max_{u,v \in V} |d_G(u, v) - d_{G'}(u, v)| \leq d \cdot d_G(u, v) + \beta$ for some constant β.

With respect to a graph G, we can also define additive and multiplicative distortion for a path P. If P is a path between $u, v \in V$, then P has an *additive distortion* of d if $\ell(P) - d_G(u, v) \leq d$, and a *multiplicative distortion* of d if $\ell(P) - d_G(u, v) \leq d \cdot d_G(u, v) + \beta$ for some constant β.

When the context (additive or multiplicative) is clear, we say that G' is a d-distortion of G if it has distortion d with respect to G. Throughout this paper, we restrict our attention to modified graphs G' which are d-distortions of the original graph G, for small constant values of d.

2.3 Queries and Responses

As stated in the introduction, one of our aims is to understand the limitations placed on watermarking schemes which have only indirect access to the data used by a suspect provider. In our setting, this indirect access is through the interface the provider offers to the end-users, returning some information $A(u, v)$ in response to a query (u, v). We consider three different models, based on the nature of the information available via the provider's interface to the end-users.

Edge Queries: If $(u, v) \in E$, $A(u, v) = \ell((u, v))$.

Distance Queries: $A(u, v) = d_G(u, v)$.

Route Queries: $A(u, v) = P$, where P is a shortest path from u to v in G (no distance information is assumed to be revealed).

Broadly speaking, the edge model gives the owner complete access to the copy of the provider. While somewhat unrealistic, it allows us to separate the limitations imposed by the small distortion requirement from the ones imposed by the indirect access, and address only the problems posed by the small distortion requirement.

Since many existing Internet map engines give both distance and route information, the ability to encode information in the more restrictive distance model has relevance to real-world situations. In examining it, we will build on techniques developed for the simpler edge model.

Finally, the route model provides the end-user, and thus the owner, with only a path connecting the query nodes, rather than distance information. Such a strategy might be employed by a cheating provider, since it allows him to give useful information (how to get from point A to point B) while suppressing additional information that may help the owner catch him. More importantly, the route model applies naturally to situations in which the provider, rather than actively providing answers to queries, instead responds by taking some action that may be observed by the owner. As an example, consider the case where the provider is a trucking company which uses maps to find shortest routes for its trucks, and the owner wishes to determine whether the trucking company makes use of his data simply by observing the routes taken by the trucks.

2.4 Adversarial vs. Nonadversarial

The next axis which we consider is whether or not the provider gives honest answers consistent with his data in response to queries. A provider using an illegally obtained copy may consider introducing additional distortion in answering queries to evade detection.

A model is *non-adversarial* if the provider answers all queries correctly, based on the copy of the map he possesses, and *adversarial* otherwise.

Clearly, in an adversarial setting, one needs to limit the behavior and resources of the provider; otherwise, he can always evade detection. Specifically, against an adversary who either provides answers with *large distortion* or has *fairly accurate* knowledge of the true map, no scheme can be effective. We rule out each of these two cases by formalizing two assumptions, namely, the *Bounded Distortion Assumption* and *Limited Knowledge Assumption*. Section 5 describes these assumptions in detail.

2.5 Marking Schemes

A *marking scheme* consists of a pair of algorithms, $\mathcal{M}$ and $\mathcal{D}$, where $\mathcal{M}$ is called the *marking algorithm* and $\mathcal{D}$ is called the *detection algorithm* (also refered to as the *detector*). A marking scheme is said to encode b bits of information with an additive (multiplicative) distortion d if the following conditions are satisfied:

1. $\mathcal{M}$ takes as input $G = \langle G, l\rangle$ and a string $r \in \{0,1\}^b$, and outputs $G_r = \langle G, l_r\rangle$ such that for any $r \in \{0,1\}^b$, G_r is a d-distortion of G.

2. $\mathcal{D}$ takes as input $\langle G, l\rangle$ and answers $\mathcal{A} = \bigcup_{u,v\in V} A(u,v)$ to every pairwise query, as provided through the end-user interface, and outputs $r \in \{0,1\}^b$.

In other words, the marking algorithm maps each provider r to a copy of the map $G_r = \langle G, l_r\rangle$ such that G_r is a d-distortion of G, and the detection algorithm uses the answers to its queries to recover the provider r to whom this copy was given.

We say that a marking scheme *has error* p if for any provider using G_r, $\Pr[\mathcal{D} \text{ outputs } r] \geq 1-p$. In the case of adversarial models, it will be necessary to consider schemes with nonzero error.

It is easy to see that if only a bounded distortion is allowed, one cannot hope to encode more than $O(m)$ bits of information. Throughout this paper, our goal is to design schemes which can encode a $\text{poly}(n,m)$ bits.

3 Overview of Results

3.1 Nonadversarial Edge and Distance Models

We start with a study of the edge and the distance models in the nonadversarial setting. These two models serve to highlight the manner in which shortest paths in a graph may change when the edge lengths are modified and reveal some inherent difficulties in watermarking maps. Our main result here is as follows:

THEOREM 3.1. *For the nonadversarial edge model, there are marking schemes that encode $\Omega(m^{1/2-\epsilon})$ bits with additive distortion and $\tilde{\Omega}(m)$ bits with multiplicative distortion. For the nonadversarial distance model, there are marking schemes that encode $\Omega(n^{1/2-\epsilon})$ bits with additive distortion and $\tilde{\Omega}(n)$ bits with multiplicative distortion. In each case, the additive distortion is $O(1/\epsilon)$ while the multiplicative distortion is $(1+o(1))$.*

3.2 Adversarial Edge and Distance Models

Building on our techniques for the nonadversarial setting above, we design a marking scheme that works in a general adversarial model. Our scheme introduces only a small additive distortion d and is effective against an adversary willing to introduce a much larger additive distortion d' in its effort to evade detection.

THEOREM 3.2. *For the adversarial edge model, there is a marking scheme that encodes $\Omega(m^{1/2-\epsilon})$ bits while for the adversarial distance model, there is a marking scheme that encodes $\Omega(n^{1/2-\epsilon})$ bits.*

3.3 Nonadversarial Route Models

Finally, we focus on marking schemes for the route model. When the queries are answered only with route information, the marking problem becomes significantly more complex. In particular, the topology of the graph as well as its length function plays a critical role. To encode information in a query (u,v) with only a small distortion, it must be the case that there is at least one *alternate* path connecting u and v that is not much longer than a shortest path. Thus effective marking schemes require not only that the topology of the underlying graph allow for multiple distinct paths between many pairs of nodes but that the underlying length function allows for multiple near-shortest paths. For instance, there are graphs G such that (i) we can encode $\Omega(n)$ bits in the route model with multiplicative distortion when all edges have similar lengths, but (ii) only $O(\log n)$ bits can be encoded when the edge lengths are allowed to be arbitrary. In a similar vein, any scheme which is allowed an additive distortion of d can not mark a clique with uniform edge lengths equal to $d+1$.

In light of such inherent limitations, design of good marking schemes in the route model requires that the input data satisfy certain properties. To begin with, we observe that the cut edges in a graph do not contribute new marking opportunities since paths between the components connected by the cut edge must use that edge. Thus, in the route model, any graph can be decomposed into a collection of 2-edge connected components without decreasing the number of bits which can be encoded in it. For this reason, we restrict our attention to 2-edge connected graphs.

We further focus our study on graphs with nearly-uniform edge lengths, using the multiplicative distortion measure.

THEOREM 3.3. *For the nonadversarial route model, there is a marking scheme that encodes $\Omega(m^{1/2-\epsilon})$ bits with small multiplicative distortion when the underlying graph is 2-edge connected and its length function is nearly-uniform.*

Rest of the paper: Due to space limitations, we present proofs only for the edge and distance model. Section 4 presents marking schemes for the nonadversarial setting while Section 5 describes our adversarial setting and develops a framework for transforming our nonadversarial marking scheme for the distance model to one for an adversarial model. We defer the proofs of our results on the route model to the full paper.

4 Nonadversarial Case

4.1 Edge Model: Additive Distortion

We now design a marking scheme that allows us to encode $O(m^{1/2-\epsilon})$ bits of information while introducing an additive distortion of $O(1/\epsilon)$. In the nonadversarial edge model, the detector can read back the suspect graph G' by examining all edges. Thus, our focus in this model is on generating many distinct marked graphs while introducing only a small distortion; once we have such a family of graphs, they can be distinguished by edge queries.

Let $P \in \mathcal{P}(G)$ be a shortest path in G with the largest number of edges, and let L denote the number of edges on P. Our marking scheme depends on whether or not L exceeds a certain threshold value L_0 (to be determined later). Let $u_0, u_1, ..., u_L$ be the nodes along the path P.

$\underline{L \geq L_0}$: In this case, our marking scheme focuses on the path P. Let X be the set of nodes u_i such that i is odd and let $Y \subseteq X$ be the set of nodes of degree two on this set. If at least a $(1/3)$-fraction of the nodes in X have degree two, we use the following procedure to create a marked copy for each $Y' \subseteq Y$. Let $u_{i_1}, ..., u_{i_k}$ be the nodes in Y'. We create a marked copy G' by modifying the length function ℓ to ℓ' in the following manner, illustrated on the left in Figure 1. For each $u_{i_j} \in Y'$, we set $\ell'((u_{i_j-1}, u_{i_j})) = \ell((u_{i_j-1}, u_{i_j})) - 1$ and $\ell'((u_{i_j}, u_{i_j+1})) = \ell((u_{i_j}, u_{i_j+1})) + 1$. The remaining edge lengths are kept unchanged.

Clearly, this scheme gives us $|Y| \geq |X|/3 \geq L_0/6$ bits. Any path P traverses incremented and decremented edges in pairs, except possibly for the first and last edges of P. From this fact, we can conclude that in this case, our scheme introduces only a small additive distortion.

PROPOSITION 4.1. *For any pair x, y of nodes, $|d_{G'}(x,y) - d_G(x,y)| \leq 2$.*

On the other hand, if at least a $(2/3)$-fraction of the nodes in X have degree at least 3, then we proceed as follows: Let F be the set of edges that are incident on some node in $X \setminus Y$ but that do not lie on P. Clearly, $|F| \geq |X \setminus Y|/2 \geq L_0/6$. For each subset $F' \subseteq F$, we create a marked copy G' by modifying the length function ℓ to ℓ' in the following manner. We increment the length of every edge in F' by 1 and keep all other edge lengths unchanged, as depicted on the right of Figure 1. This scheme gives us $|F| \geq L_0/6$ bits.

To show that this procedure introduces only small distortion, we use the following fact regarding shortest paths.

FACT 4.1. *Let P be any shortest path in a graph G. For any $x, y \in V$, there is a shortest path P' from x to y such that $P' \cap P$ is a subpath of P.*

From this fact, we can conclude

PROPOSITION 4.2. *For any pair x, y of nodes, $|d_{G'}(x,y) - d_G(x,y)| \leq 2$.*

$\underline{L \leq L_0}$: In this case, we know that every pair of nodes in G is connected by a shortest path with at most L_0 edges on it. We use this property to design an efficient marking scheme via the probabilistic method.

Given a set $E' \subseteq E$, an *E'-edge marking* of the graph G is a graph G' with a length function ℓ' defined by

$$\ell'(e) = \begin{cases} \ell(e) + 1 & \text{if } e \in E' \\ \ell(e) & \text{otherwise} \end{cases}$$

We call a set E' of edges an *ϵ-good* edge marking set if the graph G', the E'-edge marking of G, satisfies the following property:

$$d_{G'}(x,y) - d_G(x,y) < \lceil 1/\epsilon \rceil.$$

The following proposition shows that an ϵ-good edge marking set of large size can be found at random.

PROPOSITION 4.3. *Let E' be a set of edges obtained by randomly choosing each edge in E with probability $p = 1/(L_0 n^{2\epsilon})$. Then the set E' is an ϵ-good edge set of size $\Omega(pm)$, with probability at least $1/4$.*

Proof. Consider any path $P_i \in \mathcal{P}(G)$ in this collection. The probability that E' contains at least $d = \lceil 1/\epsilon \rceil$ edges from this path P_i is bounded by

$$\binom{L_i}{d} p^d \leq (L_0)^d p^d = \frac{1}{n^{2\epsilon d}} \leq \frac{1}{n^2}.$$

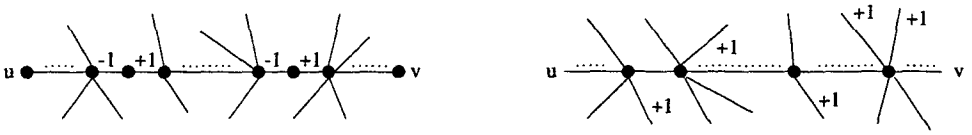

Figure 1: Edge Model: Additive Distortion

If we let G' denote an E'-edge marking of G, then for any pair u, v of nodes,

$$\Pr\left[(d_{G'}(u,v) - d_G(u,v)) \geq \lceil 1/\epsilon \rceil\right] \leq \frac{1}{n^2}.$$

Thus with probability at least 1/2, the additive distortion of G' is bounded by d. Moreover, Chernoff bounds allow us to conclude that $|E'| = \Omega(pm)$ with probability at least 3/4. The proposition then follows.

We start by constructing an ϵ-good edge marking set E^* with $\Omega(pm)$ edges. This is done by simply repeating the random experiment outlined in Proposition 4.3 and verifying that the set generated satisfies the desired properties until such a set is found. For each subset $E' \subseteq E^*$, we create a marked copy G' by taking the E'-edge marking of G. It is easy to see that we obtain $\Omega(pm)$ bits via this approach.

Our schemes for the two cases can be combined together to obtain the following result.

THEOREM 4.1. *There is a marking scheme that encodes $\Omega(m^{1/2-\epsilon})$ bits of information with only an additive distortion of $\lceil 1/\epsilon \rceil$ for any $0 < \epsilon < 1/2$.*

Proof. The marking schemes above give $\Omega(L_0)$ bits when $L \geq L_0$, and $m/(L_0 n^{2\epsilon})$ bits otherwise. Choosing the parameter L_0 to be $m^{1/2-\epsilon}$, we get the desired result.

4.2 Edge Model: Multiplicative Distortion

If we allow even a small multiplicative distortion of the form $(1 + o(1))$, we can encode $\tilde{O}(m)$ bits of information. We briefly sketch our approach here. The idea is to use the probabilistic method to construct an edge marking set that is large but introduces only a small multiplicative distortion. The important difference is that we can now pick edges with a much larger probability, say $p = 1/\log^2 n$, and use Chernoff bounds to claim that almost certainly any shortest path is stretched by a factor of $(1 + o(1))$.

THEOREM 4.2. *There is a marking scheme that gives $\tilde{O}(m)$ bits of information with only a $(1 + o(1))$-multiplicative distortion.*

4.3 Distance Model: Additive Distortion

We now consider the distance model. In moving to this setting, we must focus on the queries asked by the detector. Our strategy will be to attempt to retrieve the lengths of individual edges in G'. However, to do so we must be sure that such edges are indeed the shortest path between their endpoints even after marking the graph.

We now design a marking scheme that allows us to get $O(n^{1/2-\epsilon})$ bits of information by introducing only an additive distortion of $O(1/\epsilon)$. Our approach is similar to that presented in Section 4.1. As before, let $P \in \mathcal{P}(G)$ be a shortest path in G with a largest number of edges on it, let L denote the number of edges on P, and let $u_0, u_1, ..., u_L$ be the nodes along the path P.

$L \geq L_0$: In this case, our marking scheme focuses on the path P. Let X be the set of nodes u_i such that i is odd. We use the following procedure to create a marked copy for each $X' \subseteq X$. Let $u_{i_1}, ..., u_{i_k}$ be the nodes in X'. We create a marked copy G' by modifying the length function ℓ to ℓ' in the following manner. For each $u_{i_j} \in X'$, we set $\ell'((u_{i_j-1}, u_{i_j})) = \ell((u_{i_j-1}, u_{i_j})) - 1$ and $\ell'((u_{i_j}, v)) = \ell((u_{i_j}, v)) + 1$ for $v \neq u_{i_j-1}$. The remaining edge lengths are kept unchanged. An example is shown in Figure 2. The proposition below, which follows from Fact 4.1, shows that this scheme introduces only a small additive distortion.

PROPOSITION 4.4. *For any pair x, y of nodes, $|d_{G'}(x,y) - d_G(x,y)| \leq 3$.*

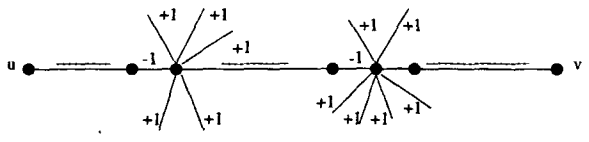

Figure 2: Distance Model: Additive Distortion

The detection algorithm is to simply query the distance between the end-points of every edge on P such that one of its end-points lie in X. Consider any edge (u, v) on the path P such that its length is decremented by 1. Since (u, v) is on a shortest path in G, $d_G(u,v) = \ell((u,v))$, and thus if we decrement

its length in some copy G', $d_{G'}(u,v) = \ell((u,v)) - 1$. Now, using the fact that our marking scheme creates a 1-1 correspondence between the marked copies and the subsets of edges on P whose length is decremented, the detection algorithm can always uniquely identify the underlying copy from the answers to its queries. Clearly, this scheme gives us $|X| \geq L_0/2$ bits.

$L \leq L_0$: We use the probabilistic method to design a marking scheme for this case. Given a set $V' \subseteq V$, a *V'-node marking* of the graph G is a graph G' with a length function ℓ' defined as follows: $\ell((u,v)) + 1$ if $u \in V'$ or $v \in V'$, and $\ell((u,v))$ otherwise.

We call a set V' of nodes an *ϵ-good* node marking set if the graph G', the V'-node marking of G, satisfies the following property: $|d_{G'}(x,y) - d_G(x,y)| < \lceil 1/\epsilon \rceil$.

The proposition below shows that an ϵ-good node marking set of a large size can be constructed at random; the proof is similar to that for edge marking sets given in Section 4.1.

PROPOSITION 4.5. *Let $V' \subseteq V$ be obtained by randomly picking each node with probability $p = 1/(L_0 n^{2\epsilon})$. Then V' is an ϵ-good node marking set of size $\Omega(pn)$ with probability at least $1/3$.*

Let V^* be an ϵ-good node marking set with $\Omega(pn)$ nodes. For each $u \in V^*$, associate an edge (u, v_u) of minimal length. We now prune V^* by constructing the graph H induced by the edges (u, v_u). In an iterative fashion, while there exists a node in H with degree at least 2, delete any such node from both H and V^*. Since each step removes at least two edges and only one node from H, this process removes at most half the nodes originally in V^*. For each subset $V' \subseteq V^*$, we create a marked copy G' by taking the V'-edge marking of G. The detection algorithm then queries (u, v_u). Since this edge is a shortest path in both G and G', the query returns $\ell'((u, v_u))$. It is easy to see that we obtain $\Omega(pn)$ bits via this approach.

Choosing $L_0 = O(n^{1/2-\epsilon})$ as the threshold, we obtain

THEOREM 4.3. *There is a marking scheme that gives $O(n^{1/2-\epsilon})$ bits of information with only an additive distortion of $\lceil 1/\epsilon \rceil$ for any $0 < \epsilon < 1/2$.*

Results for multiplicative distortion can be obtained from this construction using the techniques from Section 4.2.

5 Adversarial Case

In this section, we consider the more realistic scenario where the provider is adversarial, presenting a distorted view of his data in an attempt to either evade detection or incriminate other providers. We focus our attention on the additive distortion measure and develop a general framework that transforms our marking schemes for nonadversarial settings to ones that apply to the adversarial case. We then illustrate this framework for the distance model.

In what follows, we use $A(u,v)$ to denote the value returned by a provider of a suspect map in response to the query (u,v) where $u, v \in V$.

5.1 Model and Assumptions

As observed earlier, any reasonable model must limit the adversary's ability to distort the answers to his queries as well as the extent to which he knows about the true underlying data. We now describe these two assumptions in some detail.

The *first assumption* is easily motivated by the fact that if an adversarial provider's answers are far from reality, no one will be interested in using its services.

Assumption 1 (Bounded Distortion Assumption): For all $(u,v) \in V \times V$, $|A(u,v) - d_G(u,v)| \leq d'$, where d' is an absolute constant.

In marking the graph, we alter any edge length by at most 1, and distort distances by at most d. It is important to note that the assumption above does not require the provider's distortion to be bounded by some function of d, allowing him significant freedom to evade detections.[1]

The *second*, more subtle, limitation is that the provider has limited knowledge about the graph. From the owner's perspective, the true goal is to catch those providers who would benefit from stealing a copy of a marked graph because their own prior knowledge of the true graph is inadequate. Providers that already have detailed knowledge of the true distances independent of a copy from the owner are not relevant to us. We formalize this lack of knowledge as follows.

Given a vector $\delta \in \mathcal{Z}^E$, we define a length function ℓ_δ such that $\ell_\delta(e) = \ell(e) + \delta(e)$, and let G_δ denote the graph G with length function ℓ_δ. Now define $\Delta(u,v) = d_{G_\delta}(u,v) - d_G(u,v)$.

DEFINITION 5.1. *$W \subseteq V \times V$ is* low-bias *with respect to $S \subseteq \{0,+1,-1\}^E$ if for all $(u,v) \in W$,*

$$|\Delta(u,v)| \leq 1 \;\&\; \forall z \in \{0,+1,-1\}, \Pr_{\delta \in S}[\Delta(u,v) = z] \leq 1/2$$

[1] By contrast, if we were to restrict the provider in the same fashion as the owner, the problem would be easy: If the owner changes some distances by $\pm d$, and the provider makes any change to that graph, then unless the sign of the provider's change matches the sign of the owner's change *every* time, the resulting distortion is larger than d. However, in many situations of interest, it would be unreasonable to hold the owner and provider to exactly the same standard for distortion.

That is, for every query drawn from W, the change in distance is at most 1 and no single value predominates as we vary over S. This allows us to define sets of queries (u, v) for which $\Delta(u, v)$ is unpredictable even to someone aware of biases in $\delta(e)$ over $\delta \in S$.

DEFINITION 5.2. *$S \subseteq \{0, +1, -1\}^E$ is (γ, p)-unpredictable if for any $W \subseteq V \times V$, such that W is low-bias with respect to S and $|W| = \omega(1)$, any strategy A_{G_δ} available to the adversary satisfies*

$$\Pr_{\delta \in S}\left[\sum_{(u,v)\in W}[A_{G_\delta}(u,v) = \Delta(u,v)] > \left(\frac{1}{2}+\gamma\right)|W|\right] < p$$

Throughout, whenever we consider a probability involving the success of the strategy A, it is implicitly taken over the random choices made by A.

Assumption 2 (Limited Knowledge Assumption): For any $S \subseteq \{0, +1, -1\}^E$ such that $|S| = \omega(1)$, S is (γ, p)-unpredictable.

Since we require this assumption to hold only for large S, we allow the provider to have accurate knowledge of any constant-sized subset of the graph and instead assume only that he does not have reliable statistical information over the whole graph.

In analyzing our schemes, we will find a natural relationship between the parameters of our assumptions. A weak distortion assumption (large d') requires a strong unpredictability assumption (small γ), and vice versa.

Throughout this section d', γ, and p will refer to the parameters used in the two assumptions above.

5.2 Framework

We now present a general framework for watermarking graphs in adversarial settings. Given a graph G, our framework requires that a watermarking scheme provides a pair of distinct values $a_1, a_2 \in \{0, +1, -1\}$, a set $S \subseteq \{0, +1, -1\}^E$, and a set of queries $W = \{(u_1, v_1), \ldots, (u_L, v_L)\}$ such that

- $|W|, |S| \geq \omega(1)$.
- W is low-bias with respect to S.
- For all $\delta \in S$, $\forall (u, v) \in V \times V, |d_{G_\delta}(u, v) - d_G(u, v)| \leq d$.
- For any $D \in \{a_1, a_2\}^{|W|}$, there is a unique $\delta \in S$ such that

$$\forall i, d_{G_\delta}(u_i, v_i) = d_G(u_i, v_i) + D(i).$$

These conditions simply require the scheme to identify many independent changes which can be made consistently and do not make any reference to an adversary; thus our non-adversarial schemes fit into this framework. Let K be the number of providers; equivalently, let $\log K$ be the number of bits to be encoded. Without loss of generality, assume $a_2 > a_1$.

Marking Algorithm: For each provider r, we chose a random vector $\vec{B}^r \in \{+1, -1\}^L$. From this vector, we obtain a vector D such that $D(i) = a_2$ if $B^r(i) = +1$ and $D(i) = a_1$, otherwise. Now use D to construct δ_r as guaranteed by the framework conditions and output the graph G_{δ_r} with length function $\ell_{\delta_r} = \ell + \delta_r$.

Detection Algorithm: Given access to a suspect map, we compute an implied L-dimensional vector $\vec{Z}$, defined by $Z(i) = A(u_i, v_i)$. Let $X(i) = d_G(u_i, v_i)$ and define $a_{\text{mid}} = \frac{a_1+a_2}{2}, a_{\text{diff}} = \frac{a_2-a_1}{2}$.

For each provider r, we then compute a *similarity measure*

$$\text{sim}(\vec{B}^r, \vec{Z}) = \frac{1}{a_{\text{diff}}}\vec{B}^r \cdot (\vec{Z} - (\vec{X} + a_{\text{mid}}\vec{1}))$$

Choosing a threshold parameter $t = 0.1$, if $\text{sim}(\vec{B}^r, \vec{Z}) \geq tL$, then we say that the provider r is responsible for the suspect copy.

5.3 Analysis

By operating in this framework, we have reduced the problem of watermarking graphs to one of watermarking vectors. Associated with the graph G is a vector $\vec{X}$ which describes the correct answers to the queries in W in the original graph G. When creating a marked copy, the owner implicitly generates a vector $\vec{Y}$ with $Y(i) = X(i) + a_{\text{mid}} + a_{\text{diff}}B(i)$ representing the correct answers in the modified graph. Finally, the owner obtains a vector $\vec{Z}$ from his queries to the provider. By the Bounded Distortion Assumption, which we assume throughout this section, $\forall i, |Z(i) - X(i)| \leq d'$.

There are two possible sources of error we must consider: *false positives* (incrimination of an innocent provider) and *false negatives* (successful evasion by a guilty provider).

5.3.1 False Positives

We first show that the probability that an individual suspect provider generates a false positive is small, even for an adversary with access to the original map. We examine the $\vec{Z}$ computed by the detector from its queries to the suspect provider. First, consider the probability that it incriminates a fixed provider whose $\vec{B}$ is unknown to the suspect. Using Chernoff bounds, we can show that

PROPOSITION 5.1. *Given any valid $\vec{Z}$, $\Pr[\text{sim}(\vec{B}, \vec{Z}) > tL] \leq e^{-q^2t^2L/2}$ when $\vec{B}$ is generated randomly independent of $\vec{Z}$ and $q = \frac{a_{dif}}{d'+a_{mid}}$.*

When the detector is run, $\vec{Z}$ is compared against K such $\vec{B}$'s, one for each user. Thus, the probability that any provider is wrongly incriminated is at most $Ke^{-q^2t^2L/2}$. Since t and q are constants, we obtain the following corollary:

COROLLARY 5.1. *If $L = \Omega(\log K)$, then the probability of a false positive error by the detector is $o(1)$.*

5.3.2 False Negatives

We now show that the probability of a false negative (a guilty party evading detection) is also low. Fix a provider r; define $\vec{B} = \vec{B}^r$ and $Y(i) = \ell_{\delta_r}(e_i)$. We first show that if the provider succeeds in fooling the detector, then he has also implicitly generated a $\vec{Z}$ such that $(\vec{Z} - \vec{Y})$ is anti-correlated with $\vec{B}$.

PROPOSITION 5.2. *If $\text{sim}(\vec{B}, \vec{Z}) < tL$, then $\vec{B} \cdot (\vec{Z} - \vec{Y}) < -a_{diff}(1-t)L$.*

Then, we show that given such a $\vec{Z}$, the provider has a nonnegligible advantage in predicting $\vec{B}$ by employing a strategy of guessing $B(i)$ based on the value of $Z(i) - Y(i)$.

PROPOSITION 5.3. *Let $0 < c < d' + 1$ be a constant. Given the vector $\vec{Y}$ and a vector $\vec{Z}$ such that (i) $\vec{B} \cdot (\vec{Z} - \vec{Y}) < -cL$, and (ii) $\forall i, |Z(i) - Y(i)| \leq d' + 1$, there is an algorithm which produces a vector $\vec{C}$ such that $\sum_i [C(i) = B(i)] \geq \frac{L}{2} + \frac{cL}{4(d'+1)}$ with probability at least $1 - e^{-\Omega(L)}$.*

Proof. For each i, set

$$C(i) = \begin{cases} +1 & \text{with probability } \frac{1}{2} - \frac{Z(i)-Y(i)}{2(d'+1)} \\ -1 & \text{otherwise} \end{cases}$$

By hypothesis, this probability lies in [0,1]. Define a vector $\vec{I}$ such that $I(i)$ is the 0-1 indicator function for $C(i) = B(i)$. Note that

$$\Pr[I(i)] = \begin{cases} \frac{1}{2} - \frac{Z(i)-Y(i)}{2(d'+1)} & \text{if } B(i) = +1 \\ \frac{1}{2} + \frac{Z(i)-Y(i)}{2(d'+1)} & \text{if } B(i) = -1 \end{cases}$$

Thus,

$$\Pr[I(i)] = \frac{1}{2} - \frac{B(i)(Z(i) - Y(i))}{2(d'+1)}$$

Note that the $I(i)$'s are mutually independent, as each $I(i)$ depends only on the random choice made for $C(i)$.

Define $\alpha = \frac{c}{2(d'+1)}$, and $\mu = \mathbf{E}[\vec{I} \cdot \vec{1}] = (\frac{1}{2} + \alpha)L$.

By applying Chernoff bounds,

$$\Pr\left[\sum_i I(i) \leq (1-\gamma)\mu\right] \leq e^{-\mu\gamma^2/2}$$

Choosing $\gamma = \alpha/4$, we see that

$$(1-\gamma)\mu = (1 - \frac{\alpha}{4})(\frac{1}{2} + \alpha)L \geq (\frac{1}{2} + \frac{\alpha}{2})L$$

since $0 < \alpha \leq 1$, and that the probability of failure is at most $e^{-\Omega(L)}$.

Finally, predicting B allows the provider to predict the value of $\Delta(u, v)$ for $(u, v) \in W$. Thus, W is a counterexample to the unpredictability of S, the set of all changes made, in violation of the Limited Knowledge Assumption.

LEMMA 5.1. *If $\gamma < \frac{1}{9(d'+1)}$ and $p \geq e^{-o(L)}$, then the probability of a false negative by the detector is at most $2p$.*

Proof. The proof is by contradiction. We assume that the provider has a strategy which evades the detector with probability at least $2p$. Then taking $S = \{\delta_r\}$ over all users r and W as defined in the framework, we show that with probability greater than p, the provider has more than an γ advantage in predicting $\Delta(u, v)$, taken over all $(u, v) \in W$, contradicting the Limited Knowledge Assumption. As before, $L = |W|$.

If the provider r evades the detector, then his queries imply a vector $\vec{Z}$ such that

- $|Z(i) - X(i)| \leq d'$, by the Bounded Distortion Assumption, and
- $\vec{B}^r \cdot (\vec{Z} - \vec{Y}) < -a_{\text{diff}}(1-t)L$, since evading the detector implies $\text{sim}(\vec{B}^r, \vec{Z}) < tL$.

Given his graph $G' = G_{\delta_r}$, the provider can compute the vector $Y(i) = d_{G'}(u_i, v_i)$. Since $|a_1|, |a_2| \leq 1$, $|Z(i) - Y(i)| \leq d' + 1$. Then, he applies the algorithm from Proposition 5.3. If it succeeds, it produces a C with

$$\sum_i [C(i) = B(i)] \geq \frac{L}{2} + \frac{a_{\text{diff}}(1-t)L}{4(d'+1)}$$

By our choice of $t = 0.1$ and the fact that $a_{\text{diff}} \geq 1/2$,

$$\frac{a_{\text{diff}}(1-t)}{4(d'+1)} \geq \frac{1}{9(d'+1)} > \gamma$$

If the provider evades the detector with probability at least $2p$, and the probability that the algorithm generating C fails is $e^{-\Omega(L)}$, the provider generates such a C with probability at least p, if $p \geq e^{-o(L)}$.

Since $\Delta(u_i, v_i) = D(i) = a_{\text{mid}} + B(i)a_{\text{diff}}$, the provider's strategy $A_{G'}$ computes such a vector $C(i)$, and guesses that $\Delta(u_i, v_i) = a_{\text{mid}} + C(i)a_{\text{diff}}$. This strategy achieves

$$\Pr_{\delta \in S}\left[\sum_{(u,v)\in W} [A_{G_\delta}(u, v) = \Delta(u, v)] > \left(\frac{1}{2} + \gamma\right)|W|\right] < p$$

contradicting the (γ, p)-unpredictability of S with respect to W.

Combining this with the false positive analysis, we obtain

THEOREM 5.1. *Given a scheme consistent with the framework,* $\gamma < \frac{1}{9(d'+1)}$, *and* $p \geq e^{-o(L)}$, $O(L)$ *bits can be encoded such that the probability of error by the detector is at most* $\max\{2p, o(1)\}$.

5.4 Distance Model

We now show how this method can be used to translate our nonadversarial results to the adversarial setting. We describe this process for the additive distortion distance model, and results for the additive distortion edge model follow from similar arguments. To do so, we consider both marking schemes (for the cases $L \geq L_0$ and $L < L_0$) presented in Section 4.3. For each, we define S, W, and a_1, a_2, show that the conditions outlined in Section 5.2 are satisfied, and then apply Theorem 5.1.

If $L \geq L_0$, we define $W = \{(u_{i_j-1}, u_i) \mid u_{i_j} \in X\}$, and $a_2 = 0, a_1 = -1$. S is then the set of vectors δ obtained by taking either $\{\delta(u_{i_j-1}, u_{i_j}) = -1, \delta(u_{i_j}, u_{i_j+1}) = +1\}$ or $\{\delta(u_{i_j-1}, u_{i_j}) = 0, \delta(u_{i_j}, u_{i_j+1}) = 0\}$ independently for each $u_{i_j} \in X'$.

Choosing $L_0 = n^{1/2-\epsilon}$ as in Section 4.3, $|W|$ is $\Omega(n^{1/2-\epsilon})$, and $|S|$ is exponential in $|W|$. W is low-bias with respect to S, since each edge (u_{i_j-1}, u_{i_j}) is assigned 0 or -1 with equal probability. The maximum distortion introduced by any set of changes $\delta \in S$ is 3, by Proposition 4.4. For every $D \in \{a_1, a_2\}^{|W|}$, we set $\delta(u_{i_j-1}, u_{i_j}) = D(i)$, and $\delta(u_{i_j}, u_{i_j+1}) = -D(i)$, obtaining the unique element of S with this property. Thus, the number of bits which can be encoded is $\Omega(|W|) = \Omega(n^{1/2-\epsilon})$.

Similarly, if $L < L_0$, we choose a marking set V^* of size $O(n^{1/2-\epsilon})$ randomly and prune it as in Section 4.3. Also as before, we associate an edge v_u with each $u \in V^*$. Now, we generate the set S by associating an element $s_{V'} \in S$ (that is, $s_{V'} \in \{0, +1, -1\}^E$) with each $V' \subseteq V^*$; specifically, we take $s_{V'}((u, v)) = +1$ if $u \in V'$ or $v \in V'$ and 0 otherwise. We also choose $W = \{(u, v_u)\}$, $a_2 = +1, a_1 = 0$. Since $L_0 = n^{1/2-\epsilon}$, $|W|$ and $|S|$ are $\omega(1)$, and the arguments of Section 4.3 show that the distortion is at most $d = \lceil \frac{1}{\epsilon} \rceil$. Given $D \in \{a_1, a_2\}^{|W|}$, we obtain a unique element of S by, for all $(u, v_u) \in W$, setting $\delta(u, u') = D(i)$ for all $u' \in N(u)$. As earlier, the number of bits encoded is $\Omega(n^{1/2-\epsilon})$, and we obtain the following result.

THEOREM 5.2. *There is a marking scheme that encodes* $\Omega(n^{1/2-\epsilon})$ *bits with additive distortion* $1/\epsilon, 0 < \epsilon < 1/2$ *in the adversarial distance model.*

References

[1] ARIS Technologies. `http://www.musiccode.com`.

[2] D. Boneh and J. Shaw. Collusion secure fingerprinting for digital data. *IEEE Transactions on Information Theory*, 44(5):1897–1905, 1998.

[3] I. Cox, J. Kilian, T. Leighton, and T. Shamoon. Secure spread spectrum watermarking for multimedia. *IEEE Transactions on Image Processing*, 6:1673–1687, 1997.

[4] Digimarc. `http://www.digimarc.com`.

[5] A. B. Kahng, S. Mantik, I. L. Markov, M. Potkonjak, P. Tucker, H. Wang, and G. Wolfe. Robust IP watermarking methodologies for physical design. In *Proceedings of the 35th Annual Design Automation Conference*, pages 782–787, San Francisco, CA, June 1998.

[6] J. Kilian, F. T. Leighton, L. R. Matheson, T. G. Shamoon, R. E. Tarjan, and F. Zane. Resistance of digital fingerprints to collusional attacks. In *Proceedings of 1998 IEEE International Symposium on Information Theory*, page 271, Cambridge, MA, August 1998.

[7] A. Oliveira. Robust techniques for watermarking sequential circuit designs. In *Proceedings of the 36th Annual Design Automation Conference*, pages 837–842, New Orleans, LA, June 1999.

[8] B. Pfitzmann and M. Waidner. Anonymous fingerprinting. In *Eurocrypt '97, LNCS 1233*, pages 88–102, Berlin, 1997.

[9] B. Pfitzmann and M. Waidner. Asymmetric fingerprinting for larger collusions. In *4th ACM Conference on Computer and Communications Security*, pages 151–160, Zurich, April 1997.

[10] G. Qu and M. Potkonjak. Analysis of watermarking techniques for graph coloring problem. In *Proceedings of the 1998 IEEE/ACM International Conference on Computer Aided Design*, pages 190–193, San Jose, CA, November 1998.

[11] N. Wagner. Fingerprinting. In *Proceedings of the 1983 IEEE Symposium on Security and Privacy*, pages 18–22, April 1993.