An Information Hiding Method based on Computational Intractable Problems[⋆]

S.M. Armeni[1], D.N. Christodoulakis[1,2], I. Kostopoulos[1,2],
P.D. Kountrias[1], Y.C. Stamatiou[1,2], M. Xenos[2,3]

[1] University of Patras, Computer Engineering & Informatics Dept.,
Rio 26500, Patras, Greece.
{armeni, kostopul, pkount}@ceid.upatras.gr
http://www.ceid.upatras.gr

[2] Computer Technology Institute,
61 Riga Ferraiou Str., GR-261 10 Patras, Greece.
{dxri, stamatiu}@cti.gr
http://www.cti.gr

[3] Hellenic Open University,
Sachtouri 16, 262 23, Patras, Greece.
m.xenos@eap.gr
http://www.eap.gr

Abstract.
This paper presents an information hiding method that can be used for proving copyright ownership of a digital object. In contrast with other methods, the identity of the owner of the object is proved by the owner's ability to demonstrate *knowledge* of a certain property of the signature. The method utilizes *Zero Knowledge Interactive Proof (ZKIP)* protocols for computationally intractable, or NP-complete, problems and in particular for the 3-coloring problem. This problem requires an assignment of a color out of three available colors to the vertices of a graph so that no two adjacent vertices have the same color. The method prescribes the construction of signatures that represent *adjacency matrices* of graphs while proof of ownership is effected by knowledge of a 3-coloring by an individual. The computational intractability of the 3-coloring problem implies that knowledge of a coloring of the graph/signature presents sufficient evidence of ownership of any digital file containing this signature. The method has the additional advantage that the disclosure of the signature is of no consequence, since it is essentially the knowledge of the property of the signature (the 3-coloring of the graph it represents) that enables one to use it as proof of ownership of the cover data. The paper focuses on the design of the method and presents experiments from its application on high quality color images. From the experiments it is derived that it is feasible to add the 3-colored graph into color images without affecting image quality (PSNR measurements and HVS examples are presented). More importantly, experiments prove the signature

[⋆] Pre-print version of the paper published in *Advanves in Informatics*, Manolopoulos Y., Evripidou S., and Kakas A., (Eds.), pp. 262–278, Springer-Verlag, Vol. 2563, ISSN 0302-9743, 2003.

tolerance against various attacks even against attacks that significantly reduce image quality.

1 Introduction

It is a fact that, nowadays, digital information is more and more exposed to various acts of unauthorized modification, duplication and redistribution, not only due to the widespread use of the Internet where it appears, but also due to the invention of inexpensive data reproduction means. This has determined the emergence of a variety of signature-based protection methods (see [4, 12]) for protecting copyrighted digital material such as audio, still images, and video ensuring that the owner of the material can be identified.

The majority of these methods face two major difficulties: a) the fact that anyone can use the digital signature to claim that he is the rightful owner of the data (provided that the signature is stolen), and b) the fact that increasing the signature size either makes the method less tolerant to attacks (in case of a large non-repeatable volume of signature data), or makes the unauthorized detection and extraction of the signature easier (in case of multiple signatures). Most methods are using an embedded, user-specific bit-sequence in the material to be protected that, when detected, uniquely identifies the legal owner in cases of copyright related disputes. However, given the fact that the identification of the legal owner is effected by simply presenting the signature string, if the signature is stolen, it can be subsequently used by anyone to claim data ownership ("inversion" attacks). Additionally, most methods use a large signature consisting of repetitions of a fixed, small string consisting of different bits, usually produced randomly. In the former case, because of data repetition, the signature is easy to detect statistically. In the latter case, the unique large signature is very fragile as minor alterations of the cover data may render the signature unusable. In both cases the capacity (i.e. the ability to embed in the target image a large quantity of data) of the target image [26] is always an obstacle, since adding a large volume of data as signature may significantly decrease the quality of the image.

The method proposed in this paper, addresses both problems mentioned above. It is based on a well-known Zero Knowledge Interactive Proof (ZKIP) protocol (see Papadimitriou [24], page 291, for example) for the computationally intractable, or NP-complete, *3-coloring* problem (see [13]). The main advantage of the proposed method lies in the fact that the signature represents a graph. The legal owner of the protected data knows a 3-coloring of this graph and, thus, user identification is effected not by showing the signature/graph itself but by displaying knowledge of this 3-coloring. What is important is that such a 3-coloring is not easy to find due to the computational intractability of the 3-coloring problem. In other words, the possibility of a string to identify a person stems from its interpretation as an instance of computationally intractable problem. This has the effect that even if one steals the string, it will not be possible for him to impersonate the rightful owner of the data since a 3-coloring is needed to perform the identification.

More specifically, a user's signature represents the adjacency matrix of a random 3-colorable graph with edges to vertices ratio within a region that, experimentally, is known to produce instances that are most hard to color as shown in Section 2. In a copyright dispute, the rightful owner will be the only one able to efficiently exhibit a 3-coloring for the graph representing his signature, while the computational difficulty of producing a 3-coloring for such instances can be considered as a strong evidence in favor of him. It is for this reason that the method described here also constitutes a way for performing an alternative characterization of a strong signature. Unlike other methods, the proposed method does not focus on certain statistical signature properties (e.g. being random), or signatures representing a company logo, but on the proper representation of a combinatorial object (graph) with a hard to compute characteristic (3-coloring). In this way, digital data copyright protection is enforced, since revealing a person's digital signature (which is by itself a difficult task requiring knowledge of a, perhaps, complex embedding process) is not enough to impersonate the rightful owner; only the latter will be able to readily exhibit the 3-coloring, which is hard to compute if not already known. Such a signature, a characterization and uniqueness of which is based on a computationally intractable problem, provides an example of a cryptographically secure signature in analogy to the cryptographically secure pseudorandom number generators.

Furthermore, the proposed method uses the *wavelet transform* (see, for example, [21]), which is a transform that provides a simultaneous time-frequency representation of a given signal. Using this representation, it is possible to insert a large volume of data (a large 3-colored graph) into cover data without affecting image quality. The experiments presented in Section 5 demonstrate that the PSNR remains higher than 39dB. The Discrete Wavelet Transform is used in order to accumulate the energy of the image into the LL band. By embedding the graph information into the LL band, the algorithm achieves robustness to several image alterations since it modifies the most significant part of the image data. Considering the security of the method, two points should be considered: the robustness of the method and the security of the graph. The proposed algorithm is indeed robust to some image alterations but it cannot provide robustness against strong benchmarking tools like Stirmark or Checkmark. This does not come as a surprise since the size of information that is embedded to the image is much larger than the size of a typical watermark used for copyright purposes in most of the watermarking methods. However, even a relatively small portion of this information that remains into the cover data is still useful to the legal owner: the information bits that survive the attack define a sub-graph of the original graph and a 3-coloring of the original graph is also a 3-coloring of this sub-graph. Thus, the remaining bits still have the property that they can be used by the legal owner in order to prove ownership of the image file that contain the surviving bits.

A number of variants of the concept of Zero Knowledge authentication has been presented in the past. Namely, in [10], Craver attempts to protect the digital data by producing a scrambling that is difficult to reproduce. His scheme is based on the Hamiltonian Path problem, which is computationally intractable. Data copyright is protected by a protocol that proves ownership and at the same time avoids revealing the signature string. In another scheme proposed by Hirotsugu in [16], the topology of the signature graphs depends on (is related to) the properties of the image in which the signature graphs will be used. Hirotsugu manages to protect the copyright of an image without revealing the signature graph by taking advantage of the fact that the graph isomorphism problem is computationally intractable and by using a ZKIP protocol for it.

The approach presented in this paper differs significantly from both aforementioned approaches: even if the signature string is successfully extracted, it cannot be used unless its 3-coloring is known in advance when interpreted as a graph. The proposed Zero Knowledge variation is applicable at a second level, and is based on knowledge of a 3-coloring of a bit-string (signature) when interpreted as an instance of the 3-coloring problem. The strength of embedding instances of computationally intractable problems as watermarks in a digital image lies in the fact that even if the attacker steals the signature, he cannot use the watermark in order to impersonate the legal owner of the file. In other words, revealing the signature is of no consequences to the legal owner as a potential attacker will be in difficulty in obtaining a 3-coloring of the vertices of the graph represented by the signature. Thus, we have at the same time copyright protection due to the embedded string (watermark) and user authentication due to the knowledge of the 3-coloring of the graph represented by the string. Another difference between this approach and Hirogutsu's is that in this case the embedded graph does not depend on the cover data; this fact enables the use of the theory of threshold phenomena of hard combinatorial problems in order to construct a signature of adequate hardness.

The method described in this paper can be applied to any cover data (image, sound, video) in need of protection, although the presented experiments were conducted with still images [3]. This paper presents experiments that prove the applicability of the proposed method to high quality still color images, the ability to embed a large graph using wavelets without affecting the quality of cover data and the robustness of the method against various attacks. It should be stressed that in this case the signature's resistance against attacks does not only lie in the embedding process, but also within the properties of the graph; for this reason copyright ownership can be determined even if only a small fraction of the graph survives from the attacks (since as previously mentioned 3-coloring of it still constitutes a hard combinatorial problem).

The paper is organized as follows: Section 2 describes a method for producing random 3-colorable graphs, along with a 3-coloring of their vertices. Section 3 discusses the concept of protecting the ownership of an image by proving knowledge of the 3-coloring without revealing it, while Section 4 presents capacity issues, as well as the embedding and extracting process of the 3-colored graph. Section 5 presents results from the experiments conducted during the application of the method. Finally, in Section 6 the focal points of the paper as well as some thoughts for future investigations are discussed.

2 Producing random hard instances of the 3-coloring problem

A graph $G = (V, E)$, where V is the set of vertices and E is the set of edges of G, is 3-colorable if its vertices can be colored using at most three colors so that no two adjacent vertices are assigned the same color. A color assignment that respects this constraint is called a 3-coloring of G.

While it is difficult to color a given graph using three colors, as this problem is *computationally intractable* or *NP-complete* (see Papadimitriou [24] and Garey and Johnson [13]), it is nevertheless easy to *construct* a 3-colorable graph randomly along with a 3-coloring of its vertices. An algorithm that achieves this is the following:

1. Let p_1, p_2, and p_3 be real numbers such that $p_1+p_2+p_3 = 1$ and $p_1p_2p_3 \neq 0$.
2. Generate a random permutation $i_1, i_2, \ldots, i_n$ of the numbers $1, 2, \ldots, n$.
3. For each $j = 1, \ldots n$, vertex v_j is assigned to color class C_k with probability p_k, $k = 1, 2, 3$.
4. For each pair u, v of vertices that do not belong to the same color class, introduce the undirected edge (u, v) with probability p.

For *any* value of p, the above algorithm is guaranteed to produce a *solved* instance of the 3-coloring problem with expected color class sizes $|C_1| = p_1n$, $|C_2| = p_2n$ and $|C_3| = p_3n$. The algorithm provides both a 3-colorable graph along with a coloring of its vertices. However this does not suffice. What is actually needed, is a 3-colorable graph for which it is also hard to find a 3-coloring. In other words, a hard instance of the 3-coloring problem is sought. In general, the problem of characterizing hard instances of computationally intractable problems has led to the development of a rich theory, the theory of *instance complexity*, which is, in turn intimately related to the theory of *descriptional* or *Kolmogorov* complexity. Although the current theory does not offer an effective way to generate directly or, at least, recognize a hard instance of a computationally intractable problem, there is a number of heuristic approaches that one may follow in order to produce an instance that has some increased probability of being hard. An interesting possibility for creating such instances comes from the area of threshold phenomena in combinatorial problems.

Let G be a random graph with m edges and n vertices and r the ratio m/n. In 1991, Cheeseman, Kanefsky, and Taylor ([6]) demonstrated experimentally that for values of r that cluster around the value 2.3, randomly generated graphs with rn edges were either almost all 3-colorable or almost none 3-colorable depending on whether $r < 2.3$ or $r > 2.3$ respectively. This suggested that there is some value r_0 for r around which an abrupt transition can be observed from almost certain 3-colorability to almost certain non 3-colorability of the random graphs. However, from a complexity-theoretic perspective, their crucial observation was that graphs with edges to vertices ratio around r_0 caused the greatest difficulty to the most efficient graph coloring algorithms.

To return to the algorithm that produces solved instances of 3-colorability, it seems that the edge probability p should be selected so that the *expected* ratio $\mathbf{E}[r] = \mathbf{E}[m]/n$ is around r_0. Thus p is obtained from the following, assuming that $|C_k| = p_i n$, $k = 1, 2, 3$ (i.e. they contain exactly the expected number of vertices):

$$r_0 = \mathbf{E}[r] = \frac{\mathbf{E}[m]}{n} = p(p_1 p_2 + p_1 p_3 + p_2 p_3)n.$$

Solving for p, gives $p = \frac{r_0}{(p_1 p_2 + p_1 p_3 + p_2 p_3)n}$. It has also been experimentally observed that instances are more difficult to solve when all color classes are of about equal size so setting $p_1 = p_2 = p_3 = 1/3$ results to $p = 3r_0/n$. In this way, the "hard-instances" region is sampled.

What is not, however, clear is whether this sampling is biased towards instances that may contain many colorings besides the one produced by the proposed algorithm. This may have the undesirable effect of an opponent stealing the graph and discovering one of these colorings, thus using the graph as his own signature. A similar concern about the satisfiability problem with clauses of 3 literals (3-SAT) is expressed in [1]. The problem considered there is the generation of random formulas with clauses to variables ratio around the threshold value that are, however, constructed so as to satisfy a given truth assignment (similar to the construction of a 3-colorable graph). The problem is that the obvious methods to achieve this seem to produce formulas that, apart from the desired assignment, also satisfy many more assignments. This again might have the averse effect that when such formulas are given as inputs to randomized algorithms such as *Walksat*, they may be easier to solve than the satisfiable formulas that are sieved out of the formulas produced uniformly around the threshold value by a 3-SAT formula generator.

Another heuristic approach that can be adopted to "fortify" the instances against this deficiency is to ensure that, except for the solution which the instance is forced to have, no other solution (or, at least, a few of them) exists. Such an approach was followed in [23] for the 3-SAT problem, where it is shown that if sufficiently many clauses are chosen at random from the clauses satisfying a given truth assignment, then the formula that results possess a single solution with probability $1 - o(1)$ (see [23]). Due to the increased number of possible values per variable for 3-coloring (3 values versus 2 for 2-SAT), the technique of [23] cannot be directly generalized in the presented case. However, it can be still claimed that a number of special colorings are excluded from being legal colorings for this instance after it is constructed (with the algorithm given before): these are the colorings that result from the prescribed coloring with the change of a single color to a "higher numbered" color. Therefore, assuming that colors are numbered as 0, 1, and 2, the colorings that are excluded (see below) are the ones resulting from the initial coloring by changing the colors of single vertices to a "higher" color. More formally, a coloring P is a partition of the vertex set V into three vertex sets V_0, V_1, and V_2 so that no edge connects two vertices belonging to the same partition. A vertex u of color i is *unmovable* if every change to a higher indexed color j invalidates the coloring P. Thus, u of color i is unmovable if it is adjacent with at least one vertex of every cell V_j, such that $j > i$. A coloring P is *rigid* if all its vertices are unmovable. A good thought would, then, be to start with initial coloring that has high probability of being rigid. In [17] the probability of rigidity is computed for a given coloring P given (i) the fractions x, y, z of the chosen edges connecting vertices from V_0V_1, V_0V_2 and V_1V_2 respectively and (ii) the fractions α, β, γ of the vertices assigned color 0, 1, and 2 respectively.

$$\mathbf{P}[P \text{ is rigid given } x, y, z, \alpha, \beta, \gamma] \asymp$$
$$\left[e^{-\alpha(\int_0^1 \ln(\frac{u_{12}-x}{1-x})dx - c_{12}\ln u_{12}) - \alpha(\int_0^1 \ln(\frac{u_{13}-x}{1-x})dx - c_{13}\ln u_{13}) - \beta(\int_0^1 \ln(\frac{u_{23}-x}{1-x})dx - c_{23}\ln u_{23})}\right]^n$$

with

$$u_{12} = \frac{1}{1-e^{-c_{12}}\phi_2(c_{12}e^{-c_{12}})},\ c_{12} = \frac{xr}{\alpha}$$

$$u_{13} = \frac{1}{1-e^{-c_{13}}\phi_2(c_{13}e^{-c_{13}})},\ c_{13} = \frac{yr}{\alpha}$$

$$u_{23} = \frac{1}{1-e^{-c_{23}}\phi_2(c_{23}e^{-c_{23}})},\ c_{23} = \frac{(1-x-y)r}{\beta}$$

and $\phi_2(t)$ the smallest root of the equation $\phi_2(t) = e^{t\phi_2(t)}$. This root can be expressed using the *Lambert W function* (see [8]). The "$\asymp$" symbol means that a polynomially large factor is neglected (that can be, however, easily computed from Stirling's approximation to the factorial function obtaining an asymptotically sharp expression). As a heuristic, the values of $x, y, z, \alpha, \beta, \gamma$ can be adjusted so as to increase the estimate given above. Note that the fact that α, β, and γ enter in the probability estimate leads to the conclusion that the form of the initial coloring has an effect on how probable it is for other "neighboring" colorings to be illegal.

On the other hand, there is a trivial way to exclude many possible solutions by simply adding a large number of edges to the graph produced by the algorithm above (even making it a complete 3-partite graph). The drawback of this approach is that instances with *too few* (i.e. $o(n)$ or *too many* (i.e. $\omega(n)$) edges are easy to solve. Without delving into rigorous arguments, the former simply possess too many possible colorings. As for the latter, there is a simple randomized algorithm based on random sampling that colors a class of such graphs, called "dense" graphs, which are characterized by each vertex having $\Theta(n)$ neighbours.

Another, more rigorous and interesting approach is to construct a *certain* random class of solved 3-coloring instances with the following property: if an algorithm can be found capable of solving an instance of the class with positive (no matter how small) probability, then there is also an algorithm solving one of the NP-complete problems *in the worst case*, which would be most surprising. In [2], this goal was accomplished for a special class of *lattice problem.* We are currently trying to apply this approach for the 3-coloring problem.

What should be born in mind is that the area of hard *solved* instance generation is a very active one and there is much ongoing research aimed at building instances with a *limited* number of solutions.

3 Identification of image ownership

The application of the well known zero-knowledge interactive protocol for 3-coloring (see, e.g., [24] for details on this protocol) to signature creation and verification involves three parties:

- The owner A of the original image.
- The person B who attempts to, wrongfully, claim the image as his own.
- The referee, C.

The interactive protocol described in this paper is zero-knowledge, as far as revealing the signature to the referee is concerned, and provides a proof image ownership. The owner, A, embeds a sufficiently large, *hard to solve*, 3-colorable graph (that was produced using the procedure outlined in Section 2) in the original image I using techniques described in Section 4. What A actually embeds, is the graph's adjacency matrix which for an undirected graph without loop edges requires $n(n-1)/2$ bits. Let I' be the resulting image. This is the image that the owner A will publish (say on the Web). Suppose now that B downloads I' and claims ownership. Then A may convince the referee C that A is the owner as follows:

1. A says to C that there is a set of bits in B's image (the image I'), that represent the adjacency matrix of a graph of which we know a 3-coloring.
2. A gives to C the method to extract the positions of the bits of the adjacency matrix and A and C start executing the zero knowledge protocol for 3-coloring.

Now if B claims that he owns the image too, A challenges him to show his signature in I by telling to the referee the coordinates of the bits of his signature (I is not revealed to B). I is the *original* image which only A may possess since he only publishes the marked image I'. However, this will be impossible for him and he can only claim that A has completely removed his signature.

Therefore the referee can conclude one of the following:

1. B tells the truth and A, apart from having supposedly destroyed B's signature in I, has also managed to luckily discover some string in I' (B's "original") and then interpret it as the adjacency matrix of a graph of which A, subsequently, discovered a 3-coloring before the execution of the protocol.
2. A tells the truth, having shown a hard to compute property of a string that was embedded in I' and B has not managed to destroy this signature.

The first conclusion, if true, would lead the referee to attribute the ability to exhibit a three coloring in a graph to pure luck. However, due to the intractability of the 3-coloring problem, pure luck can be excluded with high probability. Therefore, the second conclusion is more credible than the first.

4 Data hiding process

4.1 Data hiding feasibility

There are three factors that are customarily considered in any information hiding system design: the *capacity* of the image as mentioned before, the *quality* of the new image produced after the insertion of the signature (also called watermarked image) and the *robustness* of the method. These factors are related to each other in the way shown in Figure 1 that appears in [20]. Keeping fixed one of the three factors, a trade-off relationship between the other two factors appears.

An information hiding process, generally, causes visual quality degradation, as it changes the actual values of the host data. Thus, it is important that a minimum tolerance value for a quality measure is set that will be used to evaluate the method. The proposed method has been designed so as to maintain a Peak Signal to Noise Ratio (PSNR) between the original and the watermarked image larger than 39dB. That value is considered as very high assuming the size of the signature which in this case could reach 500 Kbits (depending on the 3-colored graph and the size of the cover image).

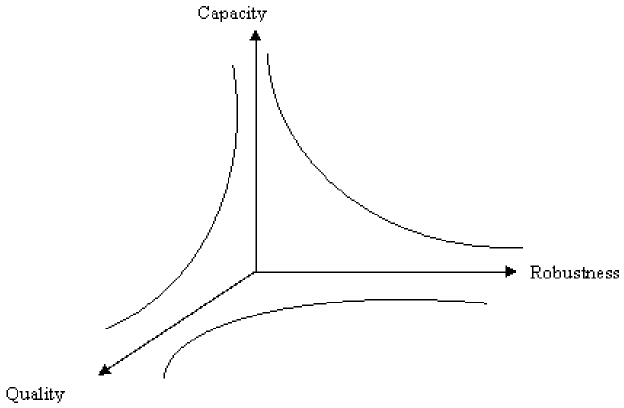

Fig. 1. Factors considered in information hiding

Since the quality of the watermarked image can be controlled to be over the tolerance value, the capacity of the image and the robustness of the method are the other two interacting factors. In the proposed method, *perceptual* evaluation criteria were chosen to evaluate the quality of the embedding of large amounts of data in digital images as well as the robustness against some of the most common attacks.

As host signals in the experiments of this work, high resolution color images with dimensions equal or larger than 1024×1024 pixels were used. The signature size for such volumes of cover data was approximately 500 Kbits, which is almost a whole bit plane after the first decomposition of the image. The embedding of the signature made use of the *discrete wavelet transform.* The wavelet transform has an advantage over other transformations, due to its ability to represent the signal in the time and frequency domain simultaneously. The signal is passed through a series of high-pass filters to analyze the high frequencies, and through a series of low-pass filters to analyze the low frequencies. In this paper, the signal is an image which can be considered as a two-dimensional signal. By applying the discrete wavelet transform in the original image, the latter is decomposed into 4 sub-images, as follows. The columns are passed through the smoothing (L) and highpass (H) filters, and the rows of each of these resultant images are passed again through L and H, this results in 4 new images. Three of them, LH, HL, and HH correspond to the highest resolution wavelet coefficients. Whereas the LL band is a smoothed version of the original and can be further decomposed in exactly the same way as the original image [25]. It should also be noted that the size of signature in this case depends on the size of the graph which is, in turn, chosen according to the size of the cover image. As a rule, for smaller cover images smaller signatures will be used and for larger images proportionately larger signatures are utilized.

The main requirement of the method is to embed the signature so that it stays intact after several attacks that could possibly applied to the watermarked image. As a consequence, the places where the signature will be embedded in the host image should be selected carefully, from the data where the most significant information of the image appears. This constraint is generally required for the design of any robust information hiding system and it is satisfied by applying a *transformation* from the spatial domain to the frequency domain (example transforms are the Discrete Cosine Transform (DCT), the Fast Fourier Transform (FFT), the Discrete Wavelet Transform (DWT) etc.).

The signature information is embedded in the LL band of the image after the first level decomposition with the discrete wavelet transform which is the most significant band. The dimensions of this band correspond to one fourth of the cover image dimensions. If, for example, a a 1024×1024 image is used, then the LL band will have 512×512 pixels. The embedding process that is described in the following section shows that hiding one bit of information in every value of the LL band is effective for embedding a large 3-colorable graph in a digital image.

4.2 Embedding the 3-colorable graph

The technique for embedding the 3-colorable graph is based on concepts presented in [5] and can, also, be applied to other kinds of digital objects such as video.

The wavelet transform is used in this technique, as it is capable of providing the time and frequency information simultaneously, giving a time-frequency representation of the signal. This kind of transform provides the opportunity to embed a large volume of data imperceptibly in the watermarked image and at the same time achieves robustness against various image filtering operations used as attacks.

A 3-colorable graph can be described by its *adjacency matrix*. Such a matrix contains zeros and ones, which represent the non-existence or existence, respectively, of edges between vertices and can be viewed as a binary (black and white) image.

The graph information is embedded into color images and its size can be the one fourth of the host image size (see Figure 2). First of all, the host image is transformed from the RGB color model into the YUV color model. The embedding process takes place in the Y component, but it can be extended using the two components U and V. Then, the 1-level discrete wavelet transform (DWT) is applied to the host image, resulting in the four sub-images LL, LH, HL and HH of the host image. The 3-colorable graph adjacency matrix bits are mapped (by means of a simple bipolar transform) from $\{0,1\}$ to $\{-1,1\}$. In addition, extra strength is given to the bipolar information by multiplying it with a factor equal to 25 that gives a good compromise between the image quality of the watermarked image and the robustness of the watermark. The graph is embedded in the LL band of the host image, using the formula $I' = I + W$, where I is the initial host image, W is the graph and I' is the final watermarked image.

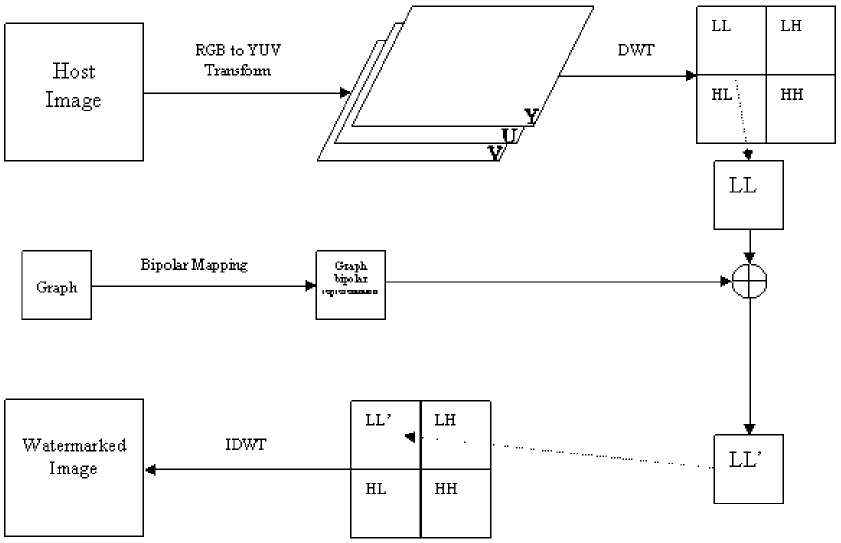

Fig. 2. The embedding process

When the embedding procedure in the LL band of the host image is completed the inverse discrete wavelet transform (IDWT) is applied and the final watermarked image is produced. Finally, this image is transformed from the YUV color model to the RGB color model.

4.3 Extracting the 3-colorable graph

In order to extract the graph, both the host and the watermarked image are needed, as shown in Figure 3. These images are, first, transformed from the RGB color model to the YUV color model. The graph information exists in the Y component of the watermarked image. Then, the 1-level discrete wavelet transform (DWT) is applied both to the Y component of the host and watermarked image. The four bands LL, LH, HL and HH are then obtained from the host and the watermarked images respectively. As mentioned before, the graph is placed in the LL band of the watermarked image. For this reason, the wavelet coefficients of the LL band of the host image are subtracted from the respective coefficients of LL band of the watermarked image to obtain the graph. If this difference is positive then the extracted graph value is equal to one. If it is negative then the extracted graph value is equal to zero.

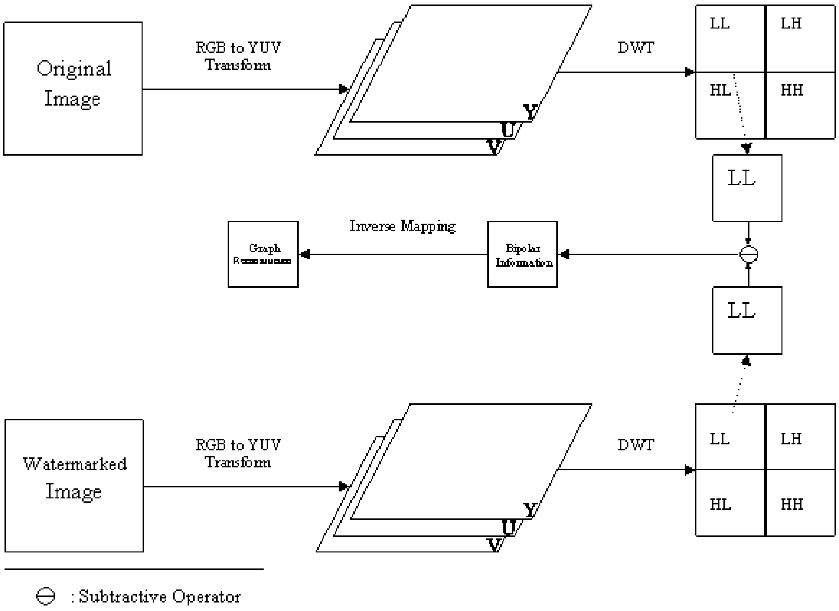

Fig. 3. The extracting process

5 Quality issues and method tolerance

Any information hiding technique inserts a volume of data into the cover data (e.g. an image). The quality of the method depends on the extent to which the original image remains unchanged which, as discussed in Section 4.1, is usually in a trade off relationship with robustness. Successful methods are those that result in the minimum possible changes to the original image. The extent to which the cover data is modified can be measured using an objective measure such as the PSNR or, subjectively, by means of the Human Visual System (HVS). The latter quality measure is based on the principle that the watermarked image should look similar to the original one to the average observer. Obviously, embedding large graphs in the original image results in small PSNR measurements while embedding smaller graphs results in higher PSNR measurements. In the method presented in this paper, the goal was to achieve a PSNR higher than 39dB. This goal is difficult to achieve as even a Jpeg Medium compression operation usually results in a PSNR measurement approximately equal to 33dB. Therefore, in order to achieve a 39dB measurement, the embedding process was designed to minimize image modifications and the size of the graph was chosen to match the size of the cover data. For an image of 1024 × 1024 pixels, it was observed that embedding a graph with over 500 vertices could result in PSNR higher than 39dB. Table 1 shows the PSNR measurements for the 6 images used in the experiments.

Image name	PSNR measurements
House	39.64
Train	39.75
Bridge	39.73
Rose	39.50
Swing	39.51
Saturn V	39.84

Table 1. PSNR-based image quality measurements

As far as the HVS criterion is concerned, no differences between the original and the watermarked images were noticeable by human observers, in any of the experiments. In Figure 4, the original image is shown on the left and the watermarked on the right.

The effectiveness of an information hiding method is also related to its robustness against modifications of the watermarked image. Such modifications, called *attacks*, can be either intentional (also called malicious, aiming at removing the embedded signature), or unintentional (i.e. a lossy compression such as Jpeg). It is obvious that the stronger the attacks are, the less tolerant a signature remains. Usually, a signature is expected to be tolerant at least to attacks affecting the watermarked image quality to a degree similar to the information hiding technique. However, the objective of the proposed method was to be tolerant to attacks that result in PSRN measurements even lower than 30dB. In order to test the robustness of the method, the following attacks were used: blurring (for evaluating robustness against attacks that amplify low frequencies), sharpening (for evaluating robustnass against attacks that amplify high frequencies), Jpeg compression using different rates (50% and 80%) and noise (adding 20% uni-

Fig. 4. Original (left) and watermarked (right) image

form noise to the watermarked image). As shown in the 6 examples of Table 2, the PSNR measurements of the watermarked images after the attacks are quite low. Blurring was always measured lower than 38dB, sharpen lower than 33dB, Jpeg-medium lower than 37dB, Jpeg-high lower than 39dB and noise lower than 27dB. As far as the 20% uniform noise is concerned, an attack that significantly affects the quality of the images, a signature is expected to have low tolerance.

Attacked Images	Blurring	Sharpen	JPEG medium	JPEG high	Uniform noise
House	36.04	31.31	33.29	38.50	26.57
Train	37.25	32.24	34.47	38.18	26.46
Bridge	36.95	32.32	36.44	39.26	26.49
Rose	36.78	32.09	33.93	38.63	26.54
Swing	36.36	31.63	33.88	38.63	26.41
Saturn V	37.07	32.30	35.51	39.19	26.59

Table 2. PSNR measurements after various attacks

A method is normally characterized as robust against attacks, if the embedded signature can be used to prove ownership of the cover data when extracted from an image altered by attacks. Table 3 illustrates the percentage of the 3-colored graph recovered for each attack case against the 6 test images.

It must be noted that in the cases of blurring, sharpening and Jpeg compression (high and medium), more than 70% of the graph was recovered in all cases (namely, in cases of blurring, sharpening and Jpeg high compression this percentage was over 80%), while in the case of noise this percentage was over 60%. As already remarked, this method is not based on the signature bits themselves to prove cover data ownership, but on the *knowledge* of a 3-coloring of the graph or sub-graphs that survived (expected to be hard themselves). Note that

Image Name	Blurring	Sharpen	JPEG medium	JPEG high	Uniform noise
House	84.12%	85.62%	71.46%	90.02%	65.71%
Train	85.11%	85.77%	79.64%	91.26%	62.92%
Bridge	92.12%	92.76%	78.04%	94.15%	65.26%
Rose	86.76%	88.10%	72.22%	90.20%	65.29%
Swing	85.67%	87.93%	74.60%	90.86%	66.39%
Saturn V	88.49%	90.03%	76.43%	91.42%	63.76%

Table 3. Percentage of adjacency matrix that survived after attacks

a 3-coloring of the original graph is also a 3-coloring for any of its subgraphs.

6 Conclusions and Future Work

Due to the advent of technological means that are able to copy and transmit information fast and at a very low cost, the need for effective methods that are able to protect copyrighted work stored in a digital form is today more apparent than ever. This need calls for solutions that are able to protect this work by identifying the legal owner in cases of copyright disputes. In this paper, a new method of protecting copyrighted work was proposed with main features (a) the shift of interest from *how user signatures look like*, as bit sequences, to *what the user signatures represent*, and (b) the use of the *wavelet transform* in order to embed signatures in the least disturbing, for the quality of the target file, way.

Regarding feature (a), the signature string is constructed so as to represent instances of hard combinatorial problems. The problem that was chosen to demonstrate the feasibility of the method was the 3-coloring problem and the signatures are adjacency matrix representations of graphs produced to possess a specific coloring that is used as a proof of identity of the legal owner. Constructing a 3-colorable random graph along with a 3-coloring of its vertices is easy while *discovering* such a coloring is hard. In addition, if these graphs are constructed within a specific region defined by a control parameter (edges to vertices ratio), the instances are expected to be considerably more difficult to solve than equal sized graphs defined outside this region. Thus, knowledge of a coloring can be used as evidence of ownership.

A major question for future research, that is also an open question in Computational Complexity Theory, is the *characterization* of hard instances. Until now, all such characterizations are either of a heuristic nature or too theoretical to be used in practice. In addition, it would be good to relate the hardness of subgraphs of a given graph to the hardness of the graph.

References

1. D. Achlioptas, C. Gomez, H. Kautz, and B. Selman, Generating satisfiable problem instances, *in Proceedings of AAAI 2000*, 2000.

2. M. Ajtai, Generating hard instances of lattice problems, Electronic Colloquium on Computational Complexity, ECCC Report TR96-007, 1996.
3. S. Armeni, D. Christodoulakis, I. Kostopoulos, Y.C. Stamatiou, and M. Xenos, Proving copyright ownership using hard instances of combinatorial intractable problems. In *Proc. 8th Panhellenic Conference in Informatics* (Nicosia, 2002), Y. Manolopoulos and S. Evripidou (eds.), 137–145, Livanis Publications, 2002.
4. W. Bender, D. Grul, N. Morimoto, and A. Lu, Techniques for data hiding, *IBM Systems Journal*, Vol. 35, NOS 3 & 4, 1996.
5. J.J. Chae, D. Mukherjee, and B.S. Manjunath, Color image embedding using multidimensional lattice structures, *in Proceedings of the IEEE International Conference on Image Processing*, Chicago, Illinois, Vol. 1, pp 460–464, October 1998.
6. P. Cheeseman, B. Kanefsky, and W.M. Taylor, Where the really hard problems are, *in Proc. of the International Joint Conference on Artificial Intelligence*, Vol. 1, pp 331–337, 1991.
7. S. Cook and D. Mitchel, Finding hard instances of the satisfiability problem: A survey, *in Satisfiability Problem: Theory and Applications*, DIMACS series in Discrte Mathematics and Theoretical Computer Science **25**, 1–17, American Mathematical Society, 1997.
8. R.M. Corless, G.H. Gonnet, D.E.G. Hare, D.J. Jeffrey, and D.E. Knuth, "On the Lambert W function," manuscript, Computer Science Department, University of Waterloo.
9. S. Craver, N. Memon, B.-L. Yeo, and M. Yeung, Can invisible watermarks resolve rightful ownerships?, *IBM Research Report*, RC 20509, 1996.
10. S. Craver, Zero knowledge watermark detection, *3rd International Workshop on Information Hiding (IHW 99)*, 1999.
11. E.R. Dougherty, Random processes for images and signal processing, *SPIE/IEEE Series on Imaging Science & Engineering*, 1999.
12. A. Fabien, P. Peticolas, R.J. Anderson, and M.G. Kuhn, Information hiding-A survey, *IEEE special issue on Protection of Multimedia Content*, 87(7), pp. 1062–1078, July 1999.
13. M.R. Garey and D.S. Johnson, *Computers and Intractability, a Guide to the Theory of NP-Completeness*, W.H. Freeman and Company, 1979.
14. R. Gonzalez and R. Woods, *Digital Image Processing*, Addison Wesley, 1992
15. B. Hayes, Computing Science: Can't get no satisfaction, *American Scientist*, March–April, 1997.
16. K. Hirotsugu, An Image Digital signature system with ZKIP for the graph isomorphism, *International Conference on Image Processing (ICIP)*, 1996.
17. A.C. Kaporis, L.M. Kirousis, and Y.C. Stamatiou, A note on the non-colorability threshold of a random graph, *Electronic Journal of Combinatorics* **7**, #R29, 2000.
18. S. Katzenbeisser, A. Fabien, and P. Petitcolas, *Information hiding techniques for steganography and digital watermarking*, Artech House Books, 1999.
19. S. Kirkpatrick and B. Selman, Critical Behavior in the satisfiability of random boolean expressions, *Science* 264, pp 1297–1301, 1994.
20. C.-Y. Lin, Watermarking and Digital Signature Techniques for Multimedia Authentication and Copyright Protection, PhD Thesis, Columbia University, 2000.
21. S. Mallat, A wavelet tour of signal processing, *Academic Press*, Second Edition, 1999.
22. F. Massacci and L. Marraro, Logical cryptanalysis as a SAT problem: the encoding of the Data Encryption Standard, manuscript, 1999.

23. M. Motoki and R. Uehara, Unique Solution Instance Generation for the 3-Satisfiability (3SAT) Problem, Technical Report C-129, Dept. of Math. and Comp. Sciences Tokyo Institute of Technology, 1999.
24. C.H. Papadimitriou, *Computational Complexity*, Addison-Wesley, 1994.
25. R. Polikar, *The Wavelet Tutorial*, Available online: http://www.polikar.iastate.edu/ rpolikar/WAVELETS/Wttutorial.html
26. K. Solanski, N. Jacobsen, S. Chandrasekaran, U. Madhow, and B.S. Manjunath, Introducing perceptual criteria into quantization based embedding, to appera in *Proc. of International Conference on Acoustics Speech and Signal Processing*, 2002.
27. R.B. Wolfgang and E.J. Delp, Overview of image security techniques with applications in multimedia systems, *in Proceedings of the SPIE Conference on Multimedia Networks: Security, Displays, Terminals, and Gateways*, Vol. 3228, pp. 297–308, Dallas, Texas, 1997.

This article was processed using the LaTeX macro package with LLNCS style