Achieving Computational and Unconditional Security in Authentication Watermarking: Analysis, Insights, and Algorithms

Chuhong Fei[a], Deepa Kundur[b], and Raymond Kwong[a]

[a]University of Toronto, 10 King's College Road, Toronto, ON Canada M5S 3G4;
[b]Taxas A&M University, 3128 TAMU, College Station, TX USA 77843-3128

ABSTRACT

This paper focuses on the analysis and enhancement of watermark-based security strategies for multimedia authentication. Based on an authentication game between a transmitter and its authorized receiver, and an opponent, security of authentication watermarking is measured by the opponent's inability to launch a successful attack. In this work, we consider two traditional classes of security for authentication: computational security and unconditional security. First we identify authentication watermarking as an error detection problem, which is different from error correction coding in robust watermarking. Then we analyze the computational and unconditional security requirements of an error detection code structure associated with quantization-based authentication watermarking schemes. We propose a novel security enhancement strategy that results in efficient and secure quantization-based embedding and verification algorithms. For computational security, cryptographic message authentication codes are incorporated while unconditional security is obtained by using unconditionally secure authentication codes. Both theoretical analysis and experimental results are presented. They show that using our approach, protection is achieved without significant increase in embedding distortion, and without sacrificing computational efficiency of the embedding and verification algorithms.

Keywords: Digital Watermarking, Authentication, Computational Security, Unconditional Security

1. INTRODUCTION

Many multimedia authentication watermarking systems have been proposed in the last few years for ensuring the integrity and origin of multimedia data such as images [1–3]. In these systems, an authenticator is imperceptibly *embedded* in the multimedia signal. The authenticator can be a sensitive function of the signal (e.g. hash) [2] or a set of coarser content features such as block histograms [4], or edge maps [5]. In comparison with appending the authenticator to the original signal, authentication watermarking systems reduce the extra storage required for the authenticator bits. Another advantage of watermark-based systems is that lossless format conversion of the secured multimedia does not necessarily change its authenticity results. Applications of authentication watermarking include trusted cameras, automatic video surveillance [6], digital insurance claim evidence [7], journalistic photography. Digital watermarking techniques are used in commercial products such as GeoVision's GV-Series digital video recorders for digital video surveillance to prevent tampering.

The goal of authentication watermarking is to authenticate the content, not its specific format representation. Thus, the embedding of the authenticator as an invisible watermark in a host signal has two main objectives: to alert a party to unacceptable distortions on the host and to authenticate the legitimate source. Possible distortions on a signal can be divided into two groups: legitimate and illegitimate distortions. When a signal undergoes a legitimate distortion which does not alter the content of the data, the authentication system should indicate that the signal is authentic. Conversely, when it undergoes illegitimate tampering, the distorted signal should be rejected as inauthentic. In contrast to robust watermarking for copyright protection applications, the watermark in authentication watermarking must survive legitimate distortions, but be fully destroyed by illegitimate modifications applied to the signal. Initially proposed digital watermarking techniques for authentication were highly fragile [1, 2] often detecting any modifications to the signal in a similar way to traditional digital signatures. In order to exploit the benefits of a data embedding approach to content authentication, *semi-fragile*

E-mail: fei@control.toronto.edu, deepa@ee.tamu.edu, kwong@control.toronto.edu

Security, Steganography, and Watermarking of Multimedia Contents VII, edited by Edward J. Delp III, Ping Wah Wong, Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 5681 © 2005 SPIE and IS&T · 0277-786X/05/$15

watermarking methods [3,8–10] were later introduced to tolerate certain kinds of processing. The primary advantage of employing semi-fragile watermarking over digital signature and fragile watermarking technology is that there is greater potential in characterizing tamper distortion, and in designing a method which is robust to certain kinds of processing. One of the first approaches to semi-fragile watermarking called telltale tamper-proofing was proposed by Kundur and Hatzinakos [11] to determine the extent of modification both in the spatial and frequency domains of a signal using a statistics-based tamper assessment function. One influential semi-fragile system is the self-authentication-and-recovery image (SARI) method developed by Lin and Chang [3] in which a semi-fragile signature is designed to survive JPEG compression up to a certain level.

Although the issues of (semi-)fragility and perceptibility are well addressed in the authentication watermarking literature [10], it has been acknowledged that the cryptographic security aspects have been neglected. Authentication "security" refers to the ability of a document or signal to resist intentional tampering by some opponent [12]. A successful multimedia authentication system must be designed to be secure against intentional tampering attacks. Compared to traditional "hard" authentication in which any slight modification to the signal is detected to be illegal tampering, the more forgiving semi-fragile systems are more vulnerable to counterfeiting and forgery [13, 14].

Therefore, to develop techniques that have the advantages of both hard and semi-fragile authentication approaches, we propose to explicitly consider security during the semi-fragile watermarking design process. Specifically, we investigate the security requirements of an authentication system by modelling the interaction between a transmitter and its authorized receiver, and an opponent as a game. In our development, we adopt *Kerckhoffs' principle* which requires that the opponent knows the details of all aspects of the authentication system except for the secret key shared between the transmitter and the receiver. We treat authentication watermarking as a communication verification problem. By employing a coding approach, we note that authentication watermarking is essentially an error detection problem, different from error correction coding employed extensively in robust watermarking. Then we analyze the security requirements of an error detection code structure associated with quantization-based authentication watermarking schemes. Two traditional classes of security for authentication are considered: computational security and unconditional security. A system is computationally secure if the opponent's effort to break it is *computationally infeasible*, while the system is defined to be unconditionally secure if it cannot be broken given infinite computational resources. Unconditional security is stronger than computational security. To satisfy the security requirements, we propose a method called MSB-LSB decomposition based on nested lattice codes, which is shown to be more secure than traditional approaches that use orthogonal domains for authenticator generation and embedding. By this approach, the security of authentication watermarking is directly associated with the security of cryptographic authentication codes. For computational security, cryptographic message authentication codes are incorporated while unconditional security is obtained by using unconditionally secure authentication codes.

The paper is organized as follows. We model authentication watermarking through the use of error detection codes and analyze the security requirements in Section 2. A novel security enhancement strategy is presented in Section 3. Unconditional security is discussed in Section 4. To demonstrate synthesis within this framework, Section 5 illustrates the design of an authentication system which is semi-fragile to JPEG compression. Conclusions are drawn in Section 6.

2. AUTHENTICATION WATERMARKING SECURITY ANALYSIS

In this section, we investigate an opponent's possible attacks on an authentication watermarking system. Then we analyze the security requirements for the watermarking scheme that the sender uses to communicate to assure integrity to the receiver.

2.1. The Authentication Game

We consider the following general symmetric key authentication model as shown in Fig. 1, which is similar to the one in [15]. The transmitter, Alice, wants to send a multimedia signal S of length n to the receiver, Bob, through a public channel. In order to facilitate analysis, we assume that the host signal S, the authenticated signal X, and the possibly distorted signal Y take values in the n-dimensional Euclidean space $\mathbb{R}^n$ (although our analysis can also be applied to other linear spaces such as the binary space $\mathbb{F}_2^n$.)

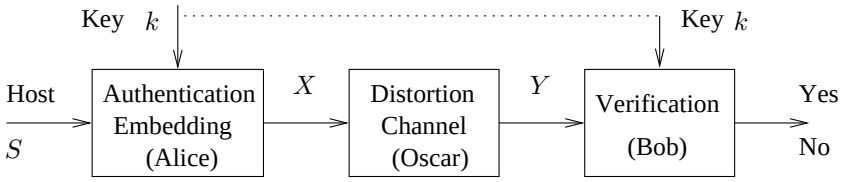

Figure 1. General authentication watermarking model.

Alice and Bob share a secret key k which can be any secret information shared between the sender and receiver. This may include the traditional key in cryptographic techniques, as well as the watermark which is kept secret for some semi-fragile authentication systems. For example, in the Yeung-Mintzer scheme [1], a watermark logo image is the secret key. In the Wong scheme [2], the secret key is an encryption key as well as a binary watermark image.

In order for Bob to be assured that the signal does originate from Alice, Alice authenticates the host source S with the secret key k to produce an authenticated signal X. The *authentication distortion* between the original signal S and the authenticated signal X, defined as $D = \frac{1}{n}E\{\|S - X\|^2\}$, should be as small as possible, so the authentication embedding is imperceptible. During transmission in the public channel, the authenticated signal X may be altered by incidental distortions, or malicious tampering by an opponent Oscar. Any modification that is applied to the authenticated signal X is modelled as the result of a distortion channel. As stated in the introduction, the distortion channel can be divided to two groups: legitimate and illegitimate. In fragile watermarking, any distortion is regarded as illegitimate while in semi-fragile watermarking, certain content-preserving distortions are considered as legitimate. Knowing the secret key k, the receiver tries to decide whether the received signal Y is authentic or not by checking if the secret key is embedded in the received signal Y. Ideally, the received signal Y is accepted as authentic with key k if the channel distortion is legitimate, and rejected as inauthentic if the channel distortion is illegitimate.

Authentication security is related to an authentication game between the sender and its authorized receiver, and the opponent in the channel. The active opponent in the channel aims to intercept the authenticated signal X, tamper or even replace it with his own signal to deceive the receiver. Therefore, the sender must use a watermarking scheme that can resist such intentional tampering to assure legitimacy. We next investigate the strategies of both parties in the authentication game.

2.1.1. Alice and Bob

From authentication theory, we know that in principle, the transmitter and the receiver only use a subset of the total number of possible messages to communicate with each other [16]. In this way, the receiver can distinguish authentic messages from fraudulent ones which are identified because they are generally not in the selected subset. Given a shared key k, we define $C(k)$ to be a set of authentic messages the transmitter uses to communicate with the receiver. Alice sends messages in this authenticated set $C(k)$ to Bob. Since Bob shares the same key k with Alice, Bob can confirm the legitimacy of messages from Alice by verifying if the received message Y is in the subset $C(k)$ associated with Alice key k. However, the opponent in the channel does not know the secret key k, so he cannot determine the subset $C(k)$. Therefore, given that $C(k)$ is a small subset of all possible messages, the opponent's probability of deceiving the receiver through random message selection is small.

From a communication and coding theory perspective, the sender and receiver in our framework employ an error detection code to detect with high probability any illegitimate tampering of the transmitted signal. This is in direct contrast to the theory of robust watermarking which makes use of error correction codes. By focusing our work on developing error detection codes that exhibit greater computational simplicity, we demonstrate that it is easier to incorporate issues of cryptographic security without sacrificing practicality.

Fig. 2 shows an example of the encoding set $\mathcal{C}(k)$. For any multimedia signal S, Alice tries to find the closest authenticated signal X in the encoding set $\mathcal{C}(k)$ that approximates the original signal, then sends the authenticated signal X to Bob. Since Bob possesses the secret key k, he has full knowledge of the encoding set $\mathcal{C}(k)$. He can therefore verify the legitimacy of the received message Y. To determine whether X has undergone

(a) Encoding Set. (b) Verification Region.

Figure 2. Encoding set and verification region for authentication watermarking. (a) All points marked with +'s form the encoding set $\mathcal{C}(k)$, which is a small subset of the signal space. (b) The shadowed area is verification region associated with key k. All codewords in the encoding set $\mathcal{C}(k)$ and their legitimate neighbors are considered as authentic.

acceptable distortion, Bob can verify if Y is in a corresponding pre-defined *verification region* as shown in Fig. 2(b). The verification region contains all codewords in the encoding set $\mathcal{C}(k)$ and their legitimate neighbors. For fragile watermarking, the verification region is equal to the encoding set $C(k)$.

The reader should note that all previously proposed authentication watermarking schemes fall into our general error detection code model. In the Yeung-Mintzer scheme [1], the encoding set $\mathcal{C}(k)$ is a set of images which map to the given binary logo image (key) by three sets of random binary lookup tables. In the Wong scheme [2], the encoding set is a set of images whose least significant bits, a hash, and a watermark bitmap satisfy an exclusive OR operation. In the SARI system [3], the encoding set is a set of images in which the authenticator derived from the authentication region matches the watermark generated from the watermarking region.

From the error detection code analysis above, we see that given an encoding set $\mathcal{C}(k)$, minimal authentication distortion is achieved using nearest neighbor embedding for the sender. Therefore, quantization-based embedding is usually employed for authentication watermarking to minimize the quantization error on average. An optimal encoding set $C(k)$ should span the signal space such that for any signal, there is an authentic codeword nearby.

2.1.2. Oscar

The opponent in the channel aims to intercept the authenticated signal that Alice sends to Bob, then possibly tamper or even replace it with his own to deceive the receiver. Oscar's attack is successful if Bob accepts as authentic the modified or fabricated signal.

In the authentication model depicted in Fig. 1, it is assumed that the opponent has full knowledge of the authentication embedding and verification details except for the secret key. The opponent can deceive the receiver in the following ways depending on the number of authentic signals he can access [16].

1. The opponent, based on his knowledge of the general authentication scheme, sends a fraudulent signal to the receiver while, in reality, the transmitter has not sent any message. Such an attack is called *impersonation*. In this attack, the opponent does not have access to an authentic message.
2. The second type of attack occurs when the opponent intercepts one or several authenticated messages from the transmitter and alters one in an *illegitimate* manner such that the modified signal is wrongly accepted as authentic by the receiver. This is called *substitution*. In this attack, the opponent has access to one or more authentic messages.

In this paper, the replay attack, in which an opponent stores and later re-transmits the authentic message, is not considered to be a valid forgery. A valid substitution must involve illegitimate tampering of an authentic message. Message replay protection can be achieved through the addition of time stamps or message indexes to a message authentication system.

The opponent's ability to successfully forge will increase as he intercepts more authenticated signals. Many specific attacks proposed in the literature to defeat authentication watermarking systems, such as vector quantization [13] and collage attacks [17], fall into the category of substitution attacks.

2.2. Security Requirements

In both attacks, the opponent will be successful if the fraudulent signal is wrongly accepted by the receiver as authentic. We know that a received signal Y is regarded as authentic from Alice if and only if Y is in the Alice's verification region. Therefore, the impersonation attack is successful if the fraudulent signal devised by the opponent happens to fall into the verification region shown in Fig. 2(b). The opponent's likelihood of randomly "choosing" a signal acceptable to the receiver, known as the probability of success of an impersonation attack P_I, is equal to the volume ratio of the verification region to the whole signal space. To reduce P_I, it is desired that the encoding set $\mathcal{C}(k)$ has as few codewords as possible. On the other hand, to reduce authentication distortion, it is desired for the encoding set to have as many codewords as possible to span the signal space. Thus, there is a subtle tradeoff between probability of success of an impersonation attack and authentication distortion.

For a substitution attack, suppose the opponent intercepts one or several authenticated signals $X \in \mathcal{C}(k)$ where X could be a sequence. He attempts to devise an illegitimate copy of X which will be wrongly accepted by the receiver. Since the opponent also knows the watermarking scheme, the substitution attack is equivalent to finding a distinct signal $X' \neq X$ such that X' is in the same encoding set $\mathcal{C}(k)$. The authentication watermarking system is secure against this type of attack if the encoding set $\mathcal{C}(k)$ has the following property: for all $k \in \mathbb{K}$, given one or several signals $X \in \mathcal{C}(k)$, it is infeasible for the opponent to find a distinct signal X' in the same encoding set $\mathcal{C}(k)$.

Security of authentication watermarking is measured by the opponent's infeasibility to launch a successful attack. This infeasibility can be evaluated by two main criteria [18]. The first measure concerns the computational efforts required for the opponent to break an authentication system. The opponent is restricted in computational resources such as the running time or memory size. A system is *computationally secure* if the best algorithm for launching a successful attack requires at least unrealistic operations or unreasonably large memory size. The second measure concerns security when there is no bound placed on the amount of computation resources available to the opponent. A system is *unconditionally secure* if it cannot be broken, even with infinite computational resources. We first focus on computational security and analyze what it requires of the encoding set for authentication watermarking. The strong security measure, unconditional security, will be investigated in Section 4.

2.2.1. Computational Security Analysis

Given an opponent has full knowledge of the authentication watermarking system except for the secret key, it must be computationally infeasible for the opponent to alter the authenticated data in an illegitimate manner such that the modified copy is wrongly accepted as legitimate. We show how several proposed semi-fragile watermark-based authentication systems fail security analysis under this stringent definition.

It is well known that many digital watermarking schemes, especially quantization-based schemes, are weak against well-designed sophisticated attacks [12]. For example, a uniform quantizer is a linear structure; that is, the integer linear combination of one or more quantizer points is also a quantizer point. This structure therefore makes it easy for an opponent to counterfeit another quantized signal from which the same watermark can be extracted. The lesson to be learned here is that uniform quantizers, or dither quantizers, cannot be used to directly construct the encoding set $\mathcal{C}(k)$ since they do not satisfy the security requirement discussed above for substitution attacks.

Furthermore, vector quantization attacks [13], and collage attacks [17] have been proposed to successfully exploit the vulnerabilities of block-wise independent watermarking schemes. Mathematically, this is because in block-wise independent watermarking schemes, the overall encoding set $\mathcal{C}(k)$ can be represented as a Cartesian product $\mathcal{C}(k_1) \bigotimes \mathcal{C}(k_2) \bigotimes \ldots \bigotimes \mathcal{C}(k_n)$, where $\mathcal{C}(k_i)$ is the encoding set for each block i and n is the total number of blocks. For this Cartesian product structure, the opponent just needs to break one block to concoct an illegitimate copy. The opponent's ability to forge is related to the weakest encoding set in all blocks instead of the overall encoding set. If the same encoding set applies to all blocks, i.e. $k_1 = k_2 = \ldots = k_n$, the opponent has access to n authentic copies for just one intercepted image, so his ability to forge a counterfeit signal increases. Therefore, caution should be exercised using block-wise structure in the design of a secure encoding set $\mathcal{C}(k)$.

In the Yeung-Mintzer scheme, a random binary lookup table (LUT) is used to specify the code structure associated with a secret key. Wu [19] generalizes the LUT generation with a constraint of allowable run of entry

bits using a Markov chain model. A lookup table $T : \mathbb{Z} \rightarrow \{0, 1\}$ is a mapping from the integer set $\mathbb{Z}$ to the binary watermark bit. The lookup table T specifies a code structure corresponding to a binary watermark. Each code set is given by $\mathcal{C}(k) = \{\lambda \in \mathbb{Z} | T(\lambda) = k\}$ for a secret binary watermark k. In order to introduce security, the lookup table is randomly generated. The randomness ensures that entries in the LUT do not leak any information about other entries in the LUT. Thus given one point $x \in \mathcal{C}(k)$ for some watermark k, the opponent cannot infer anything about other points in the same watermarked set. However, the approach of introducing uncertainty to enhance security has challenges that impede its use for authentication applications. In order for the receiver to correctly authenticate signals, the receiver must also know the lookup table. Thus, the lookup table is the secret key shared between the transmitter and the receiver. A binary lookup table of length n requires at least nH bits to transmit where H is the entropy rate of the random table, and $H = 2/3$ in the case of maximum allowable run $r = 2$ [19]. For just one 8-bit image pixel, the number of key bits required is $2^8 H \approx 170$ bits. Thus, the key size required for an entire image is unreasonably large. Another challenge is its embedding complexity. Embedding a watermark into a host signal involves searching the lookup table to find a closest point around the host signal which maps to a certain watermark. Due to the uncertainty, such searching is time-consuming, especially in a high dimensional lookup table.

Existing semi-fragile systems typically partition the source media into two disjoint regions: one for authenticator generation called the generation region, and the other for watermark embedding called the embedding region. The primary advantage of this division is that watermark embedding process does not interfere with authentication verification. For example, one can generate authenticator data from the low-frequency coefficients of the DCT blocks of an image, and embed them "interference-free" in the high-frequency coefficients of the DCT blocks. This general approach is common in semi-fragile authentication systems such as [3, 20]. The disjoint processes of authentication generation and watermark embedding are separately designed to be semi-fragile to legitimate noise. In this way, the received signal is accepted by the receiver by verifying the equality between the generated authenticator and extracted watermark. However, uniform quantizer based watermarking schemes have been applied to embed the authenticator in the embedding region. Due to the security vulnerabilities of the uniform quantizer structure discussed previously, the opponent, who has full knowledge of the embedding region, can modify the embedding region such that the same watermark is extracted. Although the watermark is typically embedded in perceptually insignificant areas, severe tampering in these areas will be intolerable. In the SARI system [3], the authenticator is derived from the block pair relationship between secretly selected DCT coefficients in the authenticator generation region. However, it is possible for the opponent to modify these coefficients illegitimately without changing their relationship. More seriously, Radhakrishnan and Memon [14] shows that the secret block pair mapping can be fully recovered by the opponent if he has access to multiple images and their authenticators using the same key.

3. SECURITY ENHANCED STRATEGY

Based on the lessons learned from previous semi-fragile authentication systems, we see that each encoding set must be securely designed. That is, given one point in the set, it must be computationally infeasible to determine any other point in the same set.

We start with a simple dither modulation QIM scheme in [21] to illustrate the idea behind our security enhancement strategy. In a binary dither modulation scheme, each code set is given by $\mathcal{C}(k) = 2\mathbb{Z} + k$ for a binary watermark secret k where $2\mathbb{Z}$ is the even integer set and the dither value is the watermark k itself. This regular structure is not secure against malicious attacks. The opponent, without knowing the watermark k but having full knowledge of the embedding scheme, can easily figure out all possible watermarked codewords in the dither quantizer $\mathcal{C}(k)$ from just one codeword. Even in the case that the quantization step size 2 is kept secret, the opponent can still determine all possible codewords if he knows just two codewords because any codeword in this dither quantizer is an integer linear combination of two distinct codewords. From this binary dither modulation example, we see that the vulnerability against malicious attacks is due to the regular structure of the code set. Thus, we assert in this paper that a secure code structure must be non-regular, yet still maintain the desired fast quantizer-based embedding and verification algorithms for practicality.

We propose a novel security enhancement strategy for quantization-based schemes by incorporating cryptographic techniques at the code level. Our idea originates from the observation that in traditional message

authentication schemes, any authenticated message is in the concatenated form of (s, a) where s is the original signal, and $a = H_k(s)$ is an appended authenticator generated from s using some keyed hash function $H_k(\cdot)$. This security mechanism can analogously apply to the binary QIM dither quantizer by viewing an even integer as the source signal and setting the dither value to be a binary authenticator generated from the even integer using a key k. The security enhanced structure is described by

$$\mathcal{C}(k) = \{i + H_k(i) | i \in 2\mathbb{Z}\} \tag{1}$$

where i is any even integer and its dither value is $H_k(i)$ for some keyed hash function. We call this security enhancement strategy the MSB-LSB decomposition approach because for any codeword in the encoding set, the least significant bit (LSB), i.e. the dither value, is an authenticator of the most significant bits (MSB).

3.1. Intuitive Explanation of Security

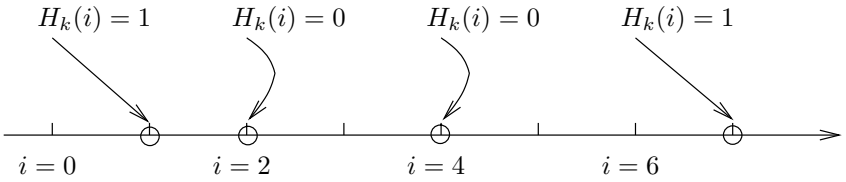

Figure 3. The offset $H_k(i)$ shifts any even number i to the points marked with $\circ$, which comprises the encoding set $\mathcal{C}(k)$.

We explain intuitively the security features of this novel encoding set $\mathcal{C}(k)$ in Eqn. (1). An example of the encoding set $\mathcal{C}(k)$ for some k and hash function $H_k(\cdot)$ is shown in Fig. 3 using the points marked with $\circ$. We can see that the encoding set is a set of integers, constructed by shifting an even integer with a corresponding dither value to a unique location. Thus the encoding set has the same cardinality as the even integer set. The dither value on an even integer i is either 0 or 1, depending on i itself as well as the key k. The authenticator generation function $H_k(i)$ is a keyed hash function, so for a fixed key k, the offset models a binary random variable. Therefore, the resulting encoding set has a non-regular structure due to pseudorandomness of the hash function. From a cryptographic point of view, the one-way hash mapping from an even integer i to a dither value m guarantees the encoding set is secure against malicious attacks. This is because without knowing the key k, it is computationally infeasible for the opponent to locate another point in the encoding set $\mathcal{C}(k)$, even if the opponent already knows one or several points in the set. In this trivial example, the authenticator is just one bit, so with probability one half, a randomly picked integer will fall into the encoding set $\mathcal{C}(k)$. However, the probability of success of impersonation will become sufficiently small when the authenticator sequence is sufficiently long.

3.2. Security Analysis of MSB-LSB enhancement strategy

We now generalize this idea to any lattice quantizer codes. We then evaluate the security of the MSB-LSB approach under the two types of attacks discussed in Section 2.1.2.

The host signal is decomposed into two components using a decomposition property of nested lattices. Consider an n-dimensional nested lattice code (Λ_1, Λ_2) where Λ_1 is the fine lattice and Λ_2 is the coarse lattice with $\Lambda_2 \subset \Lambda_1$. The fine lattice can be decomposed as $\Lambda_1 = \Lambda_2 + [\Lambda_1/\Lambda_2]$ where $[\Lambda_1/\Lambda_2]$ is the set of coset leader of Λ_1 relative to Λ_2. From this lattice decomposition, any coset lattice is represented by $\Lambda_2 + v$ where $v \in [\Lambda_1/\Lambda_2]$ is a dither vector in the coset leader set [22]. Let $H_k(\cdot)$ be the authenticator generation function mapping from the coarse quantizer Λ_2 to the coset leader set. By our MSB-LSB decomposition approach, the encoding set $\mathcal{C}(k)$ for a key k is given by $\mathcal{C}(k) = \{\lambda_2 + H_k(\lambda_2) | \lambda_2 \in \Lambda_2\}$.

For impersonation attacks, the opponent's probability of success is bounded by the volume ratio of verification region to the signal space. Since $\mathcal{C}(k)$ has the same cardinality as the coarse quantizer, P_I is equal to $1/b$ where b is the cardinality of the dither values, the range of the authenticator function $H_k(\cdot)$. Therefore, to have a small P_I, we need to use an authenticator generation function with a sufficiently large number of bits. In practice, the authenticator is at least 128 bits long, so the probability of a successful impersonation is less than 3×10^{-39}.

For the substitution attack, the opponent intercepts an authentic signal X. Let $X = \lambda_2 + v$ for some λ_2 and its authenticator v. The opponent then tries to construct an illegitimate signal $X' = \lambda_2' + v'$ such that X' will be wrongly accepted by the receiver. Since he knows all details of the method except for the key, he knows the authenticator v. The opponent can attempt to devise a signal X' from which the generated authenticator is equal to v from X. Because the authenticator is generated from the first term of the decomposition, i.e, λ_2', and the authenticator is a keyed hash, it is computationally infeasible for the opponent to find a distinct $\lambda_2' \neq \lambda_2$ such that the same authenticator v is generated as from λ_2 of X. Therefore, the opponent's feasible attack is to let λ_2' be equal to λ_2, which implies that $X' = X$, and contradicts the fact that X' is an illegitimate copy of X. Thus given one authentic signal X and a large value of n, it is computationally infeasible for the opponent to apply a substitution attack to construct such illegitimate signal X'. By this MSB-LSB decomposition approach, the computational effort for the opponent to break the encoding set $\mathcal{C}(k)$ for $k \in \mathcal{K}$ is equal to the effort to break the authenticator generation function $H_k(\cdot)$. When a keyed hash function is used for $H_k(\cdot)$, it is computationally secure against malicious attacks.

3.3. Least Distortion Authentication Embedder

The least distortion embedder creates the authenticated signal by using the nearest neighbor rule to quantize S to the closest member of the encoding set. With the structure of the encoding set $\mathcal{C}(k)$ specified in Fig. 3, we compute the distortion resulted by the security enhancement strategy. From the figure we see that in every two adjacent integers, there is one occurrence of a codeword belonging to the encoding set $\mathcal{C}(k)$. The largest distance between two adjacent codewords is 3, so the security enhancement strategy does not significantly increase the authentication distortion over using a less secure uniform quantizer. Suppose two adjacent points in the encoding set are $i + m_i$ and $i + 2 + m_{i+2}$ where i is an even integer, and m_i and m_{i+2} are two authenticators generated from some function $H_k(\cdot)$. Thus by the nearest neighbor rule, any signal between these two adjacent pair will result in a quantization noise in the range of $(-\frac{2+m_{i+2}-m_i}{2}, \frac{2+m_{i+2}-m_i}{2}]$. We assume the host signal is uniformly distributed, so the resulting quantization noise is also uniformly distributed with mean squared error $\mathrm{MSE}(m_i, m_{i+2}) = \frac{(2+m_{i+2}-m_i)^2}{12}$. If the authenticator generation function is an ideal hash function, m_i and m_{i+2} can be regarded as two independent equiprobable binary random variables. Thus, the expected authentication distortion is $D = \sum_{m_i} \sum_{m_{i+2}} \frac{1}{4}\mathrm{MSE}(m_i, m_{i+2}) = 3/8$.

To assess the significance of this result, we compare the distortion in a binary dither modulation scheme using a dither quantizer of step size 2 [21]. The embedding distortion associated with this dither quantizer is 1/3. Our proposed secure code results in distortion 12.5% more than the dither quantizer. Thus, security is achieved without significant increase in embedding distortion by our security enhancement strategy.

However, the security enhanced code now has a non-regular structure, so the nearest neighbor embedder function has to search over neighboring candidate points to compute X. The searching algorithm is time-consuming for high dimensional quantizers. However, the verification procedure is straightforward; the received signal is authentic if the extracted authenticator from the LSB part is equal to the keyed hash of the MSB component.

3.4. Suboptimal but Fast Authentication Embedder

Though the least distortion embedder is inefficient for high dimensional quantizers, it is possible to have sub-optimal but fast embedding algorithms based on quantization operations. One efficient embedding procedure decomposes a given signal S to obtain the MSB component, then shifts the MSB component by the dither value equal to the corresponding authenticator. This MSB authentication generation and LSB embedding procedure has been used in the watermarking literature. Walton [23] proposed to hide the checksums of the seven MSBs in the LSBs of pixels. We generalize this idea using a keyed hash function. Wong [2] also proposed a fragile watermarking system in which a signature is generated from the MSBs of an image block, then embedded as a watermark in the LSBs. Our implementation can be regarded as a generalization of this approach as well for semi-fragile systems that must tolerate legitimate noises.

To decompose a source signal S, we employ two quantizers: a fine quantizer with unit step size whose reconstruction point set is the integer set $\mathbb{Z}$ and a coarse quantizer with step size 2 whose reconstruction point set is the even integer set $2\mathbb{Z}$. Given a real number signal S, its quantized value λ_1 by the unit quantizer is

$\lambda_1 = Q_1(S)$ where $Q_1(S)$ is the integer rounding function. Then for the integer $\lambda_1 \in \mathbb{Z}$, the corresponding MSB and LSB components are $\lambda_2 = Q_2(\lambda_1)$ and $v = \lambda_1 \bmod 2 = \lambda_1 - Q_2(\lambda_1)$ respectively, where $Q_2(\lambda_1) = 2\lfloor \lambda_1/2 \rfloor$ is the coarse quantization function, and $\lfloor \cdot \rfloor$ is the floor function. Then we can generate a binary authenticator $m \in \{0, 1\}$ from λ_2 using a keyed hash function $H_k(\cdot)$, and embed it in the LSB component by replacing v with the derived authenticator. Thus the resulting authenticated signal is given by $X = \lambda_2 + m = \lambda_2 + H_k(\lambda_2)$.

We compute the authentication distortion induced by the above MSB and LSB embedding method. We represent the source signal as $S = \lambda_1 + r$ where $\lambda_1 \in \mathbb{Z}$ is the nearest integer around S, and the quantization noise $r \in [-0.5, 0.5)$. Then $S - X = \lambda_1 + r - (\lambda_2 + m) = v - m + r$. Again, it is assumed that the host S is uniformly distributed, so the LSB component v is an equiprobable binary distribution, and the quantization noise r is uniformly distributed in $[-0.5, 0.5)$. Suppose the authenticator generation function is ideally pseudo-random, so the authenticator m is an equiprobable binary distribution, independent of v and r. Therefore, the expected authentication distortion is given by $D = E\{|v - m + r|^2\} = Var\{v\} + Var\{m\} + Var\{r\} = \frac{1}{4} + \frac{1}{4} + \frac{1}{12} = 7/12$. Compared with the least distortion embedder of distortion 3/8, the suboptimal embedder results in more embedding distortion but achieves fast embedding algorithms.

4. UNCONDITIONAL SECURITY

We have discussed how authentication watermarking is characterized by a family of subsets, called encoding sets $\mathcal{C}(k)$ for $k \in \mathcal{K}$ where $\mathcal{K}$ is the secret key space. For computational security, elements of each encoding set $\mathcal{C}(k)$ must be computationally infeasible for an opponent to locate given that one or more are known. We have developed a security enhancement strategy to construct computationally secure encoding set. However, if the opponent is allowed to have infinite computational resources, such cryptographic enhanced codes can be broken. For example, when a digital signature scheme is used as the authenticator generation function, the encoding sets $\mathcal{C}(k)$ for different $k \in \mathcal{K}$ are disjoint. The opponent then can break the system by analyzing the secret key. Given one authenticated signal, the key can be uniquely determined by searching over key space such that the intercepted signal is a codeword of the corresponding subset. Once the secret key is determined, the opponent can send any signal at will to deceive the receiver. Thus, for unconditional security where the opponent is allowed to have infinite computational resources, the family of the encoding set $\mathcal{C}(k)$ for $k \in \mathcal{K}$ has to be well designed such that even if partial information about the key is released to the opponent, his probability of deception is still sufficiently small.

4.1. Figures of Merit

Now we formulate the opponent's best strategy to forge a signal if he has infinite computational resources. As we discussed, the code scheme for authentication watermarking is represented by a family of encoding sets $\mathcal{C}(k)$ indexed by secret key $k \in \mathcal{K}$ where $\mathcal{K}$ is the key space. We assume that the secret key k is equally distributed over the key space. The opponent has full knowledge of the code set family except for the actual key k shared by Alice and Bob. Again, there are two types of attacks: impersonation and substitution according to the number of authenticated signals he can access.

When the opponent has no access of any authenticated signal, he inserts a new signal x into the channel, the probability that x will be accepted as authentic by the receiver is the ratio of the number of keys for which x is accepted to the total number of keys, i.e. $payoff(x) = |\{k \in \mathcal{K} : x \in \mathcal{C}(k)\}|/|\mathcal{K}|$ where $|\cdot|$ denotes the cardinality of a set. The opponent will seek a signal x to maximize his chance of impersonation deception. So his maximal success probability of the impersonation attack, denoted by P_{d_0}, is $P_{d_0} = \max_x payoff(x)$.

In the substitution attack, the opponent observes an authenticated signal x and replaces it with a new signal $x' \neq x$, hoping the receiver will accept it. His probability that x' will be accepted by authentic by the receiver is the ratio of the number of keys which accept both x and x' to the number of keys that accept x, denoted by $payoff(x, x') = |\{k \in \mathcal{K} : x \in \mathcal{C}(k),\ x' \in \mathcal{C}(k)\}|/|\{k \in \mathcal{K} : x \in \mathcal{C}(k)|$. The opponent will seek a signal x' to maximize his chance of substitution deception. So the maximal success probability of the substitution attack, denoted by P_{d_1}, is $P_{d_1} = \max_x \max_{x' \neq x} payoff(x, x')$. When the opponent has access to more authenticated signals, his ability to recover the key will significantly increase. Therefore, the key in unconditional secure authentication systems is usually used only once, and a new key has to be established during each transmission

between the sender and the receiver. Under such assumption, we do not consider substitution attacks from the opponent having access to more than one authenticated signals.

It is obvious that the family of the encoding sets must be designed to deal with the worst case, which means both probabilities P_{d_0} and P_{d_1} must be as small as possible. In the following, we review unconditionally secure message authentication codes. Using the same security enhancement strategy in the previous section but by employing unconditionally secure authentication codes, we make the resulting codes $\mathcal{C}(k)$ unconditionally secure.

4.2. Authentication Code

A *systematic* authentication code (or A-code *without secrecy*) is a triple $(\mathcal{S}, \mathcal{K}, \mathcal{A})$ of finite sets together with an authentication function $f : \mathcal{S} \times \mathcal{K} \to \mathcal{A}$ [24]. Here $\mathcal{S}$ is the set of source states, $\mathcal{K}$ is the set of keys, and $\mathcal{A}$ is the set of authenticators. Every message consists of a source code with a concatenated authenticator, i.e. $m = (s, a)$ where $a = f(s, k)$. Associated with a systematic authentication code are two probabilities of success in impersonation and substitution attacks, denoted by P_I and P_S, respectively. These deception probabilities are given by

$$P_I = \max_{s,a} \frac{|\{k \in \mathcal{K} : a = f(s,k)\}|}{|\mathcal{K}|} \qquad P_S = \max_{s,a} \max_{s' \neq s, a'} \frac{|\{k \in \mathcal{K} : a = f(s,k), a' = f(s',k)\}|}{|\{k \in \mathcal{K} : a = f(s,k)\}|} \tag{2}$$

It is known that $P_I \geq 1/|\mathcal{A}|$ and $P_S \geq P_I$ [24]. The objective in the construction of authentication codes is to minimize the values P_I and P_S. Practical systematic authentication codes with small deception probabilities have been constructed using universal hashing [25], highly nonlinear functions, and algebraic curves over finite field [26].

4.3. Unconditional Security Enhancement

Similar to computational security enhancement, we can achieve unconditionally secure authentication watermarking by incorporating systematic authentication codes through the MSB-LSB decomposition. Suppose (Λ_1, Λ_2) is a nested lattice code where Λ_1 is the fine lattice and Λ_2 is the coarse lattice, and Λ_2 is a sublattice of Λ_2. Let f be a systematic authentication code on $(\Lambda_2, \mathcal{K}, [\Lambda_1/\Lambda_2])$ with deception probabilities of P_I and P_S. By the same security enhancement strategy, a family of encoding sets is constructed as follows.

$$\mathcal{C}(k) = \{\lambda_2 + f(\lambda_2, k) : \lambda_2 \in \Lambda_2\}. \tag{3}$$

It is easy to show that two deception probabilities associated with the family of encoding sets are equal to those of the incorporated authentication code. That is, $P_{d_0} = P_I$, and $P_{d_1} = P_S$. As a result, the security property of the authenticated watermarking is provided by the systematic authentication code.

5. SIMULATION RESULTS

In this section, we give a practical example to demonstrate the application of our ideas to the design of a secure semi-fragile system. The system is designed to be robust to high quality JPEG compression, but fragile to low quality JPEG compression. The design objectives of robustness and fragility are similar to the SARI system [3], but our approach also tolerates small distortions such as random Gaussian noise on DCT coefficients as well as high quality JPEG compression. Most importantly, we adopt cryptographic measures of content authentication, so security of the system is guaranteed by the incorporated message authentication codes.

JPEG compression involves quantization on DCT coefficients of an image using a quantization table specified by a compression quality factor. In DCT frequency band $(i, j), 1 \leq i, j \leq 8$, the legitimate JPEG quantization noises are bounded in the range of $(-\Delta_{ij}^q/2, \Delta_{ij}^q/2)$ where Δ_{ij}^q is the quantization step related to the pre-defined quality factor QF. For this reason, we select a fine uniform quantizer with step size Δ_{ij}^e for embedding. We set $\Delta_{ij}^e = \Delta_{ij}^q$ such that any quantization noise due to JPEG compression up to a pre-defined quality factor can be recovered correctly.

The coarse quantizer is selected according to two criteria: (1) the coset leader set from the lattice decomposition has the same size of the authenticator space; (2) the lattice is preferred to have spherical Voronoi region

to reduce the authentication distortion. One type of good lattice quantizers is the coset codes [22] constructed by concatenating a binary error correction code (ECC) with a uniform quantizer partition such as $(\mathbb{Z}/2\mathbb{Z})\Delta_{ij}^e$. For simplicity, we just apply the uniform quantizer partition $(\mathbb{Z}/2\mathbb{Z})\Delta_{ij}^e$ to some DCT coefficients to obtain least significant bits, each from one coefficient. For security, a hashed message authentication code (HMAC) based on SHA-1 is incorporated, which has a digest size of 160 bits and a key size of 256 bits. Since the security of the whole system is guaranteed in the HMAC, the location of these 160 coefficients need not be secret. These DCT coefficients are described as an ordered table $T(l) = (k, i, j), l = 1, 2, \ldots, 160$, which locates the embedding position of the l-th bit of the authenticator at the DCT coefficient $s_k(i, j)$ at the k-th block. The corresponding MSB component is those quantized values of the DCT coefficients in the table $T(l)$ by the coarse quantizer $2\mathbb{Z}\Delta_{ij}^e$, and those quantized values of the coefficients not in the table by the fine quantizer $\mathbb{Z}\Delta_{ij}^e$.

We let the pre-defined quality factor be $QF = 75$. The DCT coefficients for LSB embedding are chosen to be those in the high frequency bands for less embedding distortion. In the simulations on the test image Lenna, we notice that the integer roundoff of the pixel values after inverse DCT transform also impacts the semi-fragile system performance since the roundoff error in pixel domain is also a type of noise to DCT coefficients. We find the effect of the roundoff error in DCT domain is in the range of about $[-3, 3]$, which is is not negligible. Therefore we increase the embedding quantization step with an extra 6 to accommodate this roundoff error, so $\Delta_{ij}^e = \Delta_{ij}^q + 6$. On the test image, the modified scheme correctly outputs an authentic result when the quality factor is greater than 75, and an inauthentic result when quality factor is less than 75. The resulted authentication distortion is measured by signal-to-noise ratio (SNR) and peak signal–to-noise ratio (PSNR) with $SNR = 30.7610$ and $PSNR = 36.0818$.

We also implement an unconditionally secure scheme by incorporating a systematic authentication code constructed in [25]. The authentication code has authenticator size of 40 bits and $P_I = 2^{-40}$ and $P_S \leq 2^{-39}$. By the same MSB-LSB approach, the unconditionally secure scheme has the same ability as the HMAC scheme to distinguish high and low quality compression. On the test image, the authentication distortion is measured by $SNR = 30.9160$ $PSNR = 36.2367$, which is slightly better than the HMAC scheme since only 40 bits of authenticator need to embed in the LSB components of 40 DCT coefficients.

In summary, the features of our semi-fragile authentication scheme to JPEG compression are as follows: (1) An authenticator sequence is generated from the MSB components of the DCT coefficients using a HMAC or a systematic authentication code. Security is provided by the incorporated authenticator generation function. (2) A simple nested uniform quantizer scheme $\mathbb{Z}/2\mathbb{Z}$ makes the implementations of authentication embedding and verification easy and efficient. The authentication distortion can be further reduced by using a least distortion embedder, or by choosing an appropriate coset code for the coarse lattice. (3) Most importantly, our design framework can extend to general authentication watermarking systems. For example, our approach also applies to more complicated distortions like JPEG2000 quantization. Compromises among factors including security requirements, authentication distortion, and implementation efficiency are also readily analyzed.

6. CONCLUSIONS

In this paper, authentication watermarking is modelled as a channel error detection problem with side information to distinguish legitimate and illegitimate distortions. Security against the opponent is related to the structure of error detection codes. The results of this paper show that our proposed MSB-LSB decomposition security enhancement strategy provides useful and constructive tools to design secure quantization-based watermarking schemes for multimedia authentication. The following figures of merit are used to evaluate the security enhanced system and to compare it with other methods: authentication distortion, computational efficiency of the embedding and verification algorithms, the key space, and computational effort or probability of success for attacks by the opponent. We show that through our proposed security enhancement approach, protection is achieved without significantly increasing embedding distortion, and without sacrificing computational efficiency of the embedding and verification algorithms. Simulations on real images demonstrate the effectiveness of the security enhancement strategy.

REFERENCES

1. M. M. Yeung and F. Mintzer, "An invisible watermarking technique for image verification," in *Proc. IEEE Int. Conf. on Image Processing*, (Santa Barbara, CA), Oct. 1997.
2. P. W. Wong, "A public key watermark for image verification and authentication," in *Proc. IEEE Int. Conf. on Image Processing*, **I**, May 98.
3. C.-Y. Lin and S.-F. Chang, "Semi-fragile watermarking for authenticating JPEG visual content," in *Proc. SPIE: Security and Watermarking of Multimedia Content II*, Jan. 2000.
4. M. Schneider and S.-F. Chang, "A robust content based digital signature for image authentication," in *Proc. IEEE Int. Conf. on Image Processing*, **3**, pp. 227–230, 1996.
5. J. Dittmann, A. Steinmetz, and R. Steinmetz, "Content-based digital signature for motion pictures authentication and content-fragile watermarking," in *Proc. IEEE Int. Conf. Multimedia Computing and Systems*, **2**, pp. 209–213, 1999.
6. F. Bartolini, A. Tefas, M. Barni, and I. Pitas, "Image authentication techniques for surveillance applications," *Proc. IEEE* **89**, pp. 1403–1418, Oct. 2001.
7. K. Toyokawa, N. Morimoto, S. Tonegawa, K. Kamijo, and A. Koide, "Secure digital photograph handling with watermarking technique in insurance claim process," in *Proc. SPIE*, **3971**, pp. 438–445, 2000.
8. E. T. Lin, C. I. Podilchuk, and E. J. Delp, "Detection of image alterations using semi-fragile watermarks," in *Proc. SPIE*, **3971**, pp. 152–163, 2000.
9. C. Fei, D. Kundur, and R. Kwong, "Analysis and design of authentication watermarking," in *Proc. SPIE: Security, Steganography, and Watermarking of Multimedia Content VI*, **5306**, (San Jose, CA), Jan. 2004.
10. B. B. Zhu, M. D. Swanson, and A. H. Tewfik, "When seeing isn't believing," *IEEE Signal Processing Mag.* **21**, Mar. 2004.
11. D. Kundur and D. Hatzinakos, "Digital watermarking for telltale tamper-proofing and authentication," *Proc. IEEE* **87**, pp. 1167–1180, July 1999.
12. I. J. Cox, M. L. Miller, and J. A. Bloom, *Digital watermarking*, Morgan Kaufmann Publishers, 2002.
13. M. Holliman and N. Memon, "Counterfeiting attacks on oblivious block-wise independent invisible watermarking schemes," *IEEE Trans. Image Processing* **9**, pp. 432–441, Mar. 2000.
14. R. Radhakrishnan and N. Memon, "On the security of digest function in the SARI image authentication system," *IEEE Trans. Circuits Syst. Video Technol.* **12**, pp. 1030–1033, Nov. 2002.
15. N. Memon, P. Vora, B.-L. Yeo, and M. Yeung, "Distortion bounded authentication techniques," *Proc. SPIE* **3971**, pp. 164–174, Jan. 2000.
16. G. J. Simmons, "A survey of infomation integrity," in *Contemporary Cryptology: The Science of Information Integrity*, G. J. Simmons, ed., IEEE Press: New York, 1992.
17. J. Fridrich, M. Goljan, and N. Memon, "Further attacks on Yeung-Mintzer fragile watermarking scheme," in *Proc. SPIE: Security and Watermarking of Multimedia Contents*, pp. 428–437, (San Jose, CA), 2000.
18. D. R. Stinson, *Cryptography Theory and Practice*, Chapman & Hall/CRC, 2 ed., 2002.
19. M. Wu, "Joint security and robustness enhancement for quantization based data hiding," *IEEE Trans. Circuits Syst. Video Technol.* **13**, pp. 831–841, Aug. 2003.
20. Y. Zhao, P. Campisi, and D. Kundur, "Dual domain watermarking for authentication and compression of cultural heritage images," *IEEE Trans. Image Processing* **13**, Feb. 2004.
21. B. Chen and G. W. Wornell, "Quantization index modulation: a class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Inform. Theory* **47**, pp. 1423–1443, May 2001.
22. G. D. Forney, Jr., "Coset codes - Part I: Introduction and geometrical classification," *IEEE Trans. Inform. Theory* **34**, pp. 1123–1151, Sept. 1988.
23. S. Walton, "Image authentication for a slippery new age," *Dr. Dobb's Journal* **20**, pp. 18–26, Apr. 1995.
24. D. R. Stinson, "Combinatorial characterizations of authentication codes," *Design, Codes and Cryptography* **2**, pp. 175–187, 1992.
25. D. R. Stinson, "Universal hasing and authentication codes," *Design, Codes and Cryptography* **4**, pp. 369–380, 1994.
26. C. Xing, H. Wang, and K. Y. Lam, "Constructions of authentication codes from algebraic curves over finite fields," *IEEE Trans. Inform. Theory* **46**, pp. 886–892, May 2000.