Security Analysis of Scalar Costa Scheme Against Known Message Attack in DCT-Domain Image Watermarking

Reza Samadi*
Department of Electrical Engineering, Ferdowsi University of Mashhad, Mashhad, Iran
r.samadi@stu.um.ac.ir
Seyed Alireza Seyedin
Department of Electrical Engineering, Ferdowsi University of Mashhad, Mashhad, Iran
seyedin@um.ac.ir

Received: 08/Dec/2013 Revised: 26/May/2014 Accepted: 14/Jun/2014

Abstract

This paper proposes an accurate information-theoretic security analysis of Scalar Costa Scheme (SCS) when the SCS is employed in the embedding layer of digital image watermarking. For this purpose, Discrete Cosine Transform (DCT) coefficients are extracted from the cover images. Then, the SCS is used to embed watermarking messages into mid-frequency DCT coefficients. To prevent unauthorized embedding and/or decoding, the SCS codebook is randomized using the pseudorandom dither signal which plays the role of the secret key. A passive attacker applies Known Message Attack (KMA) on the watermarked messages to practically estimate the secret key. The security level is measured using residual entropy (equivocation) of the secret key provided that the attacker's observations are available. It can be seen that the practical security level of the SCS depends on the host statistics which has contradiction with previous theoretical result. Furthermore, the practical security analysis of the SCS leads to the different values of the residual entropy in comparison with previous theoretical equation. It will be shown that these differences are mainly due to existence of uniform regions in images that cannot be captured by previous theoretical analysis. Another source of such differences is ignoring the dependencies between the observations of non-uniform regions in previous theoretical analysis. To provide an accurate reformulation, the theoretical equation for the uniform regions and the empirical equation for the non-uniform regions are proposed. Then, by combining these equations a new equation is presented for the whole image which considers both host statistics and observations dependencies. Finally, accuracy of the proposed formulations is examined through exhaustive simulations.

Keywords: Scalar Costa Scheme; Known Message Attack; Discrete Cosine Transform; Residual Entropy; Watermarking.

1. Introduction

Digital watermarking refers to hiding a verification message into cover or host data in a secure and robust manner. In digital watermarking, the message is encoded to the watermark using a key, and then the watermark is embedded or added into the host. In the receiver side, the message should be estimated without having host signal which is called blind decoding; hence, the host plays the role of the interference in the receiver. Perfect host rejection (interference cancellation) is possible only in Ideal Costa Scheme (ICS) [1] which is optimal but non-practical. Scalar Costa Scheme (SCS) [2] is the down-to-earth implementation of the ICS which approximates the random codebook of the ICS by a set of scalar quantizers. The SCS can be used in the embedding layer in the spatial or transform domain of the image, audio, video or any other documents. The essential assumption on the quantization-based data hiding analysis is *flat-host assumption*. It assumes that the host probability density function (pdf) is uniform in each quantization cell [3]. This is equivalent to infinite Document to Watermark Ratio (DWR).

Watermarking system designers always need to have accurate closed form equation of security and robustness to reliably design such systems. An important characteristic is security level which has been poorly investigated in the literature. In the watermarking security, the purpose of the attacker is to disclose the secret parameters to implement tampering attack. Watermarking security deals with intentional attacks whose aim are accessing the watermarking channel [4], excluding those already encompassed in the robustness category. Such access refers to trying to remove, detect and estimate, write and modify the raw watermarking messages. According to the type of the information the attacker has access to, there are three scenarios, including Known Message Attack (KMA), Watermark Only Attack (WOA) and Known Original Attack (KOA) [5]. Also, there are two information-theoretic [6] and probabilistic [7] frameworks to evaluate the security level of the watermarking techniques. Information-theoretic security analysis of the SCS based on the flat-host assumption has been carried out theoretically in [8], [9]. In these studies, the authors proved that the security level of the SCS is independent of the host statistics when flat-host assumption is valid or the DWR is infinite. These

* Corresponding Author

analyzes are rather general and should be applicable in any spatial/transform domain for any kind of the cover data of image/audio/video/text. However these theoretical results may not be valid in practice due to the implementation considerations. For example these analyses cannot capture the effect of the host statistics. More precisely, work of [10] shows that breaking the SCS is practically feasible with the aid of key guessing and depends on the host statistics. Also, more recent work of [11] shows that the practical security level of the SCS in image watermarking application is much lower than what has been previously derived in [8] and strictly depends on the host statistics.

In this paper, we propose an accurate security analysis of the SCS for the digital image watermarking in Discrete Cosine Transform (DCT) domain. We apply security attack on the SCS in the KMA scenario, and evaluate the information-theoretic security level practically. It is shown in this paper that the image uniformity invalidates the flat-host assumption. Hence, the previous theoretical formulation in [8], derived based on the flat-host assumption, would be no longer valid for the uniform regions. It should be noted that each image is composed of uniform and non-uniform regions. Non-uniform regions do not invalidate the flat-host assumption; hence, it is expected that the results of [8] are valid for the non-uniform regions at least. However, we show there are dependencies between the observations that can be used by the attacker, which have not been considered by the theoretical formulation in [8]. Hence, the previous theoretical formulation in [8] would also not be valid for the non-uniform regions. To provide an accurate reformulation, we propose a theoretical equation for the uniform regions and an empirical equation for the non-uniform regions which considers host statistics and dependencies between the observations. Then, we combine our result to propose a new unified closed-form equation for the security level of the SCS in the image watermarking application. These accurate analyses are crucial when the SCS is used in the complex scenarios or real applications which need high security [12].

The remaining of this paper is organized as follows. Section 2 presents the block diagram, notations and primary definitions of the watermarking in the DCT domain using the SCS encoding. In Section 3, the practical key estimator of the SCS is reviewed and applied to the KMA scenario [8] in the DCT domain and then, the security level stressing is evaluated and compared to the former theoretical result. The theoretical-empirical equation for the security level of the SCS in KMA scenario in the DCT domain image watermarking is presented in Section 4. Finally, Section 5 concludes the paper.

2. Watermarking Using SCS in DCT domain

Block diagram of the watermarking in the DCT domain [13] is shown in Fig. 1. First, the image pixels are subdivided into 8×8 blocks and the DCT is applied on each block. Then, each 8×8 block of the DCT coefficients is ordered in a zigzag mode and the appropriate (mid-frequency) coefficients are selected for embedding. The Illustration of the zigzag ordering and selected mid-frequency DCT coefficients is presented in Fig. 2. In the embedding layer, messages M_j: $j=1,..,M$ are embedded in the selected DCT coefficients (host) X_i: $i=1,...,N$ by using the secret key K. Embedding can be done redundantly in such a way that the whole image is covered. In simple repetition coding and embedding, each message M_j is embedded in $L=|N/M|$ different hosts ,X_i , resulting the watermarked host Y_i as:

$$Y_i=\text{embedd}(X_i,M_j,K), \quad (j - 1)L<i\leq jL \tag{1}$$

After the embedding, both non-selected and selected coefficients are merged and reordered to make 8×8 blocks in the transformed domain. Then, the inverse DCT is applied on each block to create 8×8 blocks in the spatial domain. Finally, these blocks are merged, yielding the watermarked image. This image is open access and ready to travel over the network. We exclude the robustness attacks and any security mechanism other than the secret key. Under the Kerckhoffs' assumption [14], all the parameters and details of the data hiding scheme are known to the attacker, except the secret key. Hence, the attacker conducts the feature extraction on the watermarked image to access the watermarked host.

2.1 SCS for Information Embedding

In this paper, the theoretical model of the information embedding layer between the sender, proposed in [8], is employed. This model is shown in Fig. 3, where the host signal X is independent and identically distributed (i.i.d) scalar feature extracted from the original digital content. The embedder hides an equiprobable message $M \in \{0,\dots,p-1\}$ in the host by using secret key K yielding watermark W. Then, the watermark is added or embedded into the host, producing the watermarked host Y. The Attacker has access to the watermarked host through the attack channel which produces attacked host Z. Detector receives the attacked host and tries to estimates the embedded message. However, only the legal detector has the secret shared key and is able to estimate the embedded message.

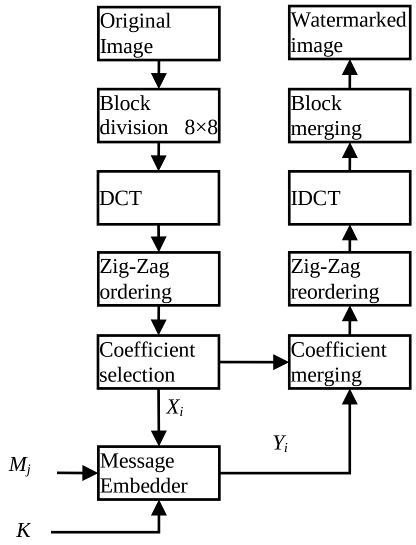


Fig. 1. Block diagram of watermarking in DCT domain

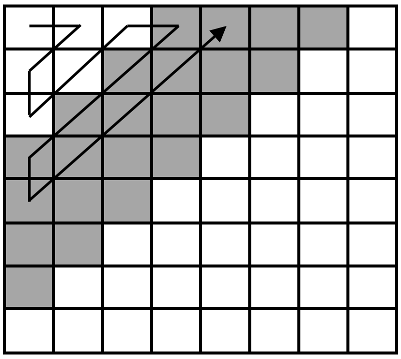


Fig. 2. Illustration of Zig-Zag ordering and selected mid-frequency DCT coefficients of Fig. 1.

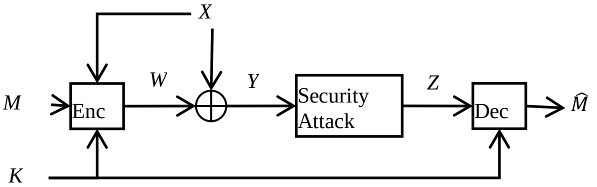


Fig. 3. Theoretical model for additive side-informed watermarking, including security attack and excluding robustness attack/noisy channel

In the SCS, watermarked host in (1) is read as (2), where $e_{M,K}(X)$ is the quantization noise of host over the shifted lattice [15], Q_Δ indicates the uniform scalar quantizer over Voronoi region $[-\Delta/2, \Delta/2]$, Δ is quantization step size and M represents the message symbol to be transmitted. The quantization noise of the host over the lattice Λ is defined as $X \mathrm{mod} \Lambda = X - Q_\Delta(X)$.

$$Y = X - \alpha e_{M,K}(X) = (1-\alpha) e_{M,K}(X) + Q_{M,K}(X) \qquad (2)$$

$$e_{M,K}(X) = X - \Delta\frac{M}{p} - K - Q_\Delta(X - \Delta\frac{M}{p} - K) \qquad (3)$$

The security of the embedder relies only on the randomization of the codebook via the secret key. The fundamental assumption when analyzing the SCS is that the secret key is distributed uniformly over the Voronoi region. Hence, the error signal $e_{M,K}(X)$ is orthogonal to X,M and the host rejection is possible [2]. In the remainder of the paper, the uniform secret key assumption is employed. The embedding distortion (watermark power) is evaluated as $D_w = \alpha^2\Delta^2/12$. Moreover, the transparency is measured by the host variance to the watermark power ratio $\lambda = \sigma_X^2/D_w$ and DWR is defined as $DWR = 10\log_{10}(\lambda)$.

3. Practical Security Level Evaluation of SCS in KMA Scenario

This paper concentrates on the KMA scenario only, where the attacker has access to the pool of observations with independent messages and corresponding watermarked hosts, all watermarked with the *same* secret key. Attacker's purpose is to disclose the secret key with the aid of the available observations and then perform the tampering attack. The security level is evaluated as the effort of the attacker to reduce his ambiguity about the secret key. In the information-theoretic framework, γ-security level is defined as the number of the N_o observations attacker needs to make $h(K|Y^{N_o}, M^{N_o})\gamma$ [8]. To evaluate the information-theoretic security level of the SCS in the KMA scenario, the residual entropy $h(K|Y^{N_o}, M^{N_o})$ is computed for every N_o. Based on the definitions presented above, the residual entropy has been analyzed theoretically in [8] for $\alpha \geq 0.5$ using flat-host assumption, as follow:

$$h_o(K|Y^{N_o}, M^{N_o}) = log_2(1-\alpha)\,\Delta + 1 - \sum_{i=1}^{N_o} \frac{1}{i} \qquad (4)$$

The subscript 'o' indicates (4) is the previous theoretical equation in comparison with the proposed equation in Section 4. In [8], the authors assumed that residual entropy is independent of the host statistics for high DWR. Detailed analysis of the security attack in [8] is reviewed in Section 3.1. In Section 3.2, we practically evaluate the residual entropy of the SCS for information embedding in the DCT domain for some test images with various host statistics. To do this, first the KMA is applied on the SCS, and then the security level is evaluated. The test images and distribution of their selected DCT coefficients (host) are shown in Fig. 4. To compare the shape of hosts' pdf only, variances of the hosts' pdf are normalized to $\sigma_x^2 = 1$.

3.1 KMA on SCS

The best practical key estimator of the SCS in the KMA scenario was proposed in [8], for $\alpha \geq 0.5$ based on the k-means clustering. Feasible value of the secret key in observation index r is derived in (5) where the set $\mathcal{Z}(\Lambda) = \{x: |x| \leq (1-\alpha)\Delta/2\}$ is the scaled version of the Voronoi region. Feasible values of the secret key after N_o observations are the intersection of the sets $\mathcal{D}_r$ as follow.

$$k \in \mathcal{D}_r,\quad \mathcal{D}_r = \left(y_r - \Delta\frac{m_r}{p} - \mathcal{Z}(\Lambda)\right) mod\Lambda \tag{5}$$

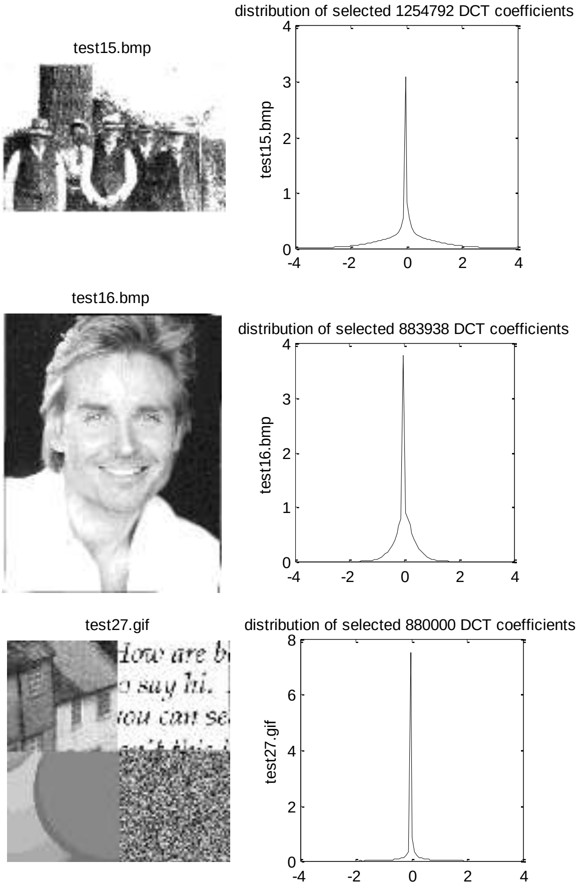
$$\mathcal{S}_{N_o} \triangleq \bigcap\nolimits_{r=1}^{N_o} \mathcal{D}_r \tag{6}$$

When the flat-host assumption is valid, the secret key is uniformly distributed on feasible region $\mathcal{S}_{\mathrm{N_o}}$ after $\mathrm{N_o}$ observations. Hence, the residual entropy of the secret key, when the watermarked host and message are provided, can be written as (7).

$$h(K|Y^{N_o} = y^{N_o}, M^{N_o} = m^{N_o}) = E[log(vol(\mathcal{S}_{N_o}))] \tag{7}$$

3.2 Implementation of KMA on SCS and Practical Security Level Evaluation

In this paper, any robustness attack is excluded. Hence, the attacker has access to the noiseless watermarked hosts and corresponding messages. Since there are *N* different hosts, hence we can *independently* implement the KMA on the SCS $\mathrm{Q} = \lfloor \mathrm{N}/\mathrm{N_o} \rfloor$ times in different trials. Then, the accurate approximation of residual entropy is computed by averaging over the trials. We define $\mathcal{S}_{\mathrm{N_o}}^{\mathrm{q}}$ as the feasible set of the secret keys in the q-th trial of the practical entropy computation based on the $\mathrm{N_o}$ observations. The definitions of the feasible sets $\mathcal{D}_\mathrm{r}$ and $\mathcal{S}_{\mathrm{N_o}}^{\mathrm{q}}$ in the model used in this paper are given in (8) and (9), respectively. Finally, the practical residual entropy can be approximated by (10).



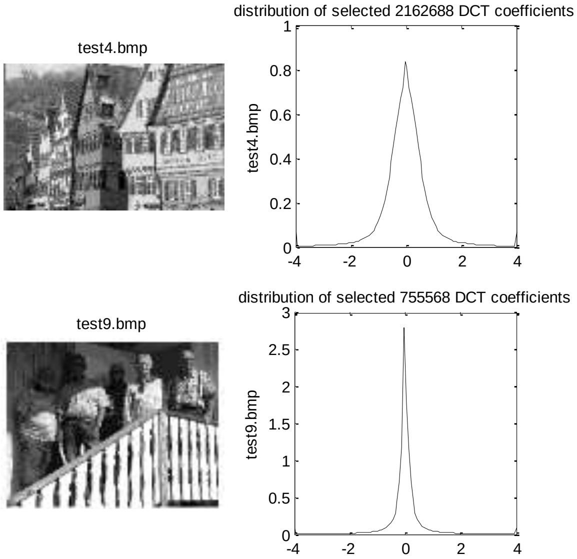


Fig. 4. Used test images and their selected DCT coefficients distribution. In each image, the variance of the coefficients is normalized to 1. Notice the differences in the shape of the distributions.

$$\mathcal{D}_r = \left(y_r - \Delta\frac{m_s}{p} - \mathcal{Z}(\Lambda)\right) mod\Lambda,\quad s = \lceil r/L \rceil \tag{8}$$

$$\mathcal{S}_{N_o}^{q} \triangleq \bigcap_{r\in((q-1)N_o,qN_o]} \mathcal{D}_r \tag{9}$$

$$h_p(K|Y^{N_o}, M^{N_o}) \cong \frac{1}{Q}\sum\nolimits_{q=1}^{Q} log(vol(\mathcal{S}_{N_o}^{q})) \tag{10}$$

The subscript ‘p’ in (10) denotes the practical residual entropy equation. Approximation (10) is resulted by application of the *weak law of large numbers* [16]. As *Q* increases, the variance of the approximation error tends to zero if the observation sequences, with the length N_0, are independent from each other for each *q*, which is a valid assumption in our problem. The proof directly follows from the assumption that the watermarked hosts y_r are function of the i.i.d random variables x_r,m_r, which are mutually independent. Moreover, for each *q* the secret key *k* is constant.

The residual entropy in (10) is evaluated for the test images of Fig. 4. These results are shown in Fig. 5, compared to the theoretical results for the SCS [8] and Spread Spectrum (SS) schemes [17]. For the sake of the fair comparison, all the test images have embedded with the same embedding parameters, $\sigma_x, \alpha, \mathrm{DWR}$. It is worth noting that all test images only differ in the shape of the host pdf. It is obvious that the change the host pdf has large impact on the residual entropy, indicating the strong dependency of the security level of the SCS in the KMA scenario on the host statistics.

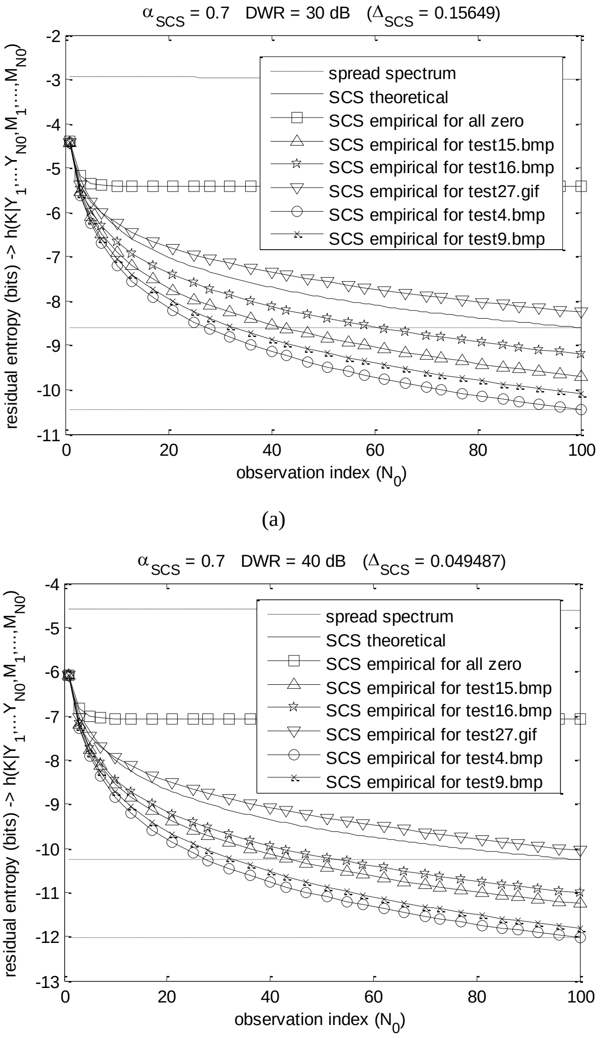
To clearly show the differences between the empirical and theoretical results, the security level is numerically computed in

Table. 1. We use those values of γ which lead to the security level of N_o≈*100* in the theoretical result for both *DWR=30,40dB*. It is easy to note the large gap between the empirical and theoretical evaluations of security level.

For example, previous result suggests $N_o \approx 100$ observations are required for $DWR=40dB$ to reduce the residual entropy lower than $\gamma = -8.6$. However, it can be seen that the practical KMA on the "test4.bmp" only needs $N_o=27$ observations to achieve this goal. Therefore, the theoretical result overestimates the security level of the SCS against the KMA and ignores the effect of the shape of the hosts' pdf on it, too.

3.3 on the need of proposing an accurate analysis

The results of this section clearly demonstrate the large gap between the practical evaluation and theoretical result in [8] for the security analysis of the SCS in the DCT domain and the KMA scenario. Theoretical result in [8] overestimates the security level and also ignores the impact of host statistics on it. Therefore, it is necessary to propose an accurate equation for the security level of the SCS.



(a)

(b)

Fig. 5. Residual entropy of SCS embedding in DCT domain in KMA scenario for five test images and specified parameters; comparisons are done with theoretical results of SCS and Spread Spectrum. The case of "all zero" is corresponds to uniform (homochromatic) image.

Table. 1. Security level of test images for two values of DWR, γ is chosen in such a way that theoretical result gives security level as 100

Image	Security level No	
	DWR=30dB $\gamma = -8.6$	DWR=40dB $\gamma = -10.26$
theoretical	100	100
test4.bmp	27	27
test9.bmp	32	32
test15.bmp	41	44
test16.bmp	60	52
test27.gif	> 100	> 100
all zero	+∞	+∞

4. Security Analysis of SCS in KMA Scenario

In previous section it was shown that the security levels of the SCS used in DCT domain for different images in the KMA scenario are not identical even for the same embedding parameters. The main goal of this section is to provide an accurate closed-form equation for this security level. To do this, each image is divided into two textured and uniform regions. We prove that the security level for the uniform regions is *constant* for different images and is not a function of the observation index. Also, it will be demonstrated that the security levels of the textured regions are *identical* for different images and are a function of the observation index. We propose a theoretical equation for the security level of the uniform regions and an empirical equation for the security level of the textured regions. Then, these results are combined to derive a unified theoretical-empirical equation for every image. Finally, the accuracy of the proposed equations is verified using a huge set of different test images.

4.1 Security analysis for uniform regions

Uniform regions of each image are adjacent pixels with low brightness variations. The DCT transform of these regions are composed of almost zeros coefficients in low frequencies in the transform domain. For example, the test image *"test27.gif"* in Fig. 4 has much uniform regions. It is clear that for the uniform region, the flat host assumption is not valid anymore. Hence the equation (4) would be no longer valid, too. We prove that the security level of the SCS for uniform regions is stated as follows:

$$h_u(K|Y^{N_o}, M^{N_o}) = log_2(1-\alpha)\,\Delta - 1 \qquad (11)$$

where, subscript 'u' the proposed theoretical equation in (11) applied to the uniform regions. The proof of (11) is straightforward: Attacker's objective is to estimate the secret key k after $\mathrm{N_o}$ observations by evaluating the set $\mathcal{S}_{\mathrm{N_o}}$ in (6). Given $\{t_i,m_i\}$, the set $\mathcal{D}_\mathrm{i}$ in (5) can be rewritten as follow:

$$\mathcal{D}_i = \left((y_i - \Delta \frac{m_i}{p}) mod\Lambda - \mathcal{Z}(\Lambda) \right) mod\Lambda, \qquad (12)$$
$$i = 1, \dots, N_o$$

The first step by the attacker is modulo lattice reduction of the observations as $\tilde{Y}_i = (Y_i - \Delta M_i/p) mod\Lambda$. Under the flat-host assumption, the attacker does not lose any information about the secret key after this modulo reduction. By substituting y_i from (2), the modulo lattice reduced observations is given by:

$$\tilde{Y}_i = ((1-\alpha)e_{M_i,K}(X_i) + K) mod\Lambda \qquad (13)$$

It is clear that the secret key is only concealed by the embedding message and quantization error $e_{m_i,K}(x_i)$. Here, the quantization error plays the role of the host interference. For the uniform regions, the selected mid-frequency coefficients are almost zero $x_i \cong 0$. Hence, after some simplifications, the quantization error for the binary embedding *(p=2)* can be evaluated as follow:

$$e_{m_i,K}(x_i) = \begin{cases} -k, & m_i = 0 \\ \frac{\Delta}{2} - k, & m_i = 1 \end{cases} \qquad (14)$$

Substituting (14) into (13) and then resulting equation into (12), will yield the following:

$$\mathcal{D}_i = \begin{cases} (k\alpha - \mathcal{Z}(\Lambda)) mod\Lambda & m_i = 0 \\ (k\alpha + \frac{(1-\alpha)\Delta}{2} - \mathcal{Z}(\Lambda)) mod\Lambda & m_i = 1 \end{cases} \qquad (15)$$

For $m_i = 0$, the set $\mathcal{D}_i$ is centered on the resized secret key with maximum ambiguity $(1-\alpha)\Delta$. On the other hand, for $m_i = 1$, the set $\mathcal{D}_i$ is centered on shifted and resized secret key with maximum ambiguity $(1-\alpha)\Delta$. Intersection of these sets gives the estimated key. However, the sets corresponding to the identical embedding message are identical. Hence, the intersection of just two sets with different embedding messages produces the same result as the intersection of all available sets. Two sets with the different embedding messages are illustrated in Fig. 6. It is easy to see that the intersection of these sets gives the intended result in (11).

4.2 Security analysis for textured regions

Non-uniform regions are called textured regions. These regions are produced by removing the uniform regions from an image. This can be realized by removing nearly zero DCT coefficients in the transform domain. It will be shown that the security level for the textured regions is identical for different images.

After removing nearly zero DCT coefficients, the security levels of the images in Fig. 4 are sketched in Fig. 7. It is interesting to note that the security levels for different textured images are identical for the same embedding parameters. Moreover, it is expected that the security level for textured region should be consistent with the previous theoretical equation of (4). However, as it can be seen in Fig. 7, the practical security level is much lower than that of the previous theoretical equation. This happens because (4) ignores the dependencies between the observations watermarked with a same secret key. Precisely, authors of [8] assumed that the sets $\mathcal{D}_i$ are independent during the proof of their theorem in [8-Appendix A]. However, from (12) and (13) we can rewrite $\mathcal{D}_i$ as:

$$\mathcal{D}_i = (K + (1-\alpha)e_{M_i,K}(X_i) - \mathcal{Z}(\Lambda)) mod\Lambda \qquad (16)$$

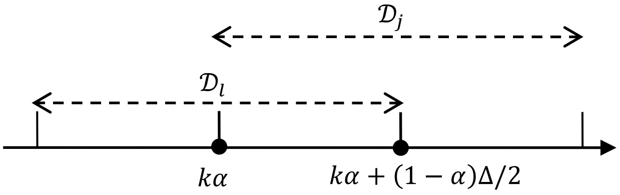
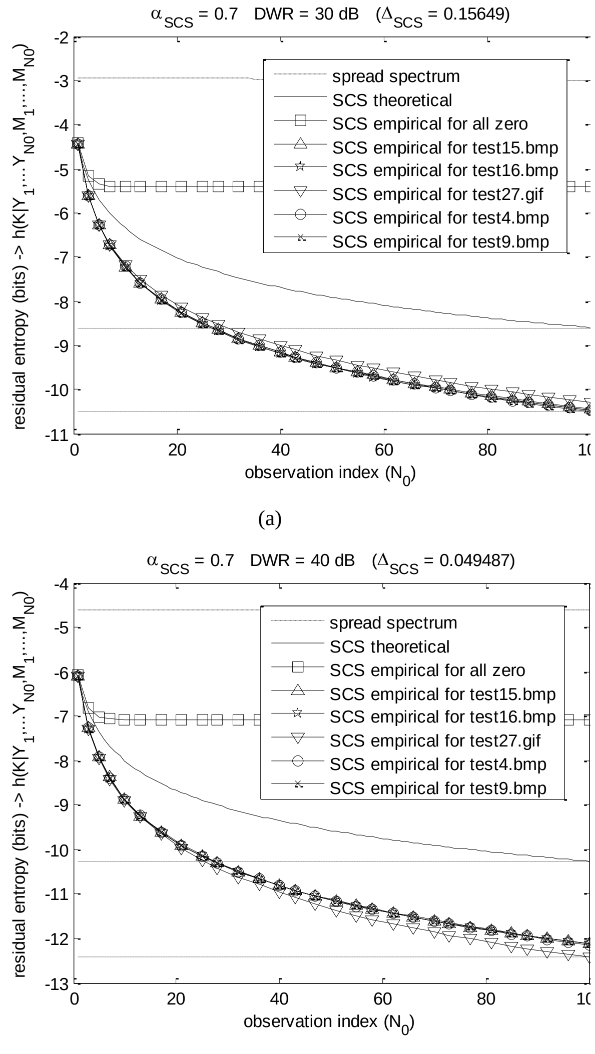


Fig. 6. Illustration of two sets of estimated secret keys for uniform regions with different embedding messages



(a)

(b)

Fig. 7. Residual entropy of SCS embedding in DCT domain in KMA scenario for five test images and specified parameters; uniform regions of test images are removed; comparisons are done with theoretical results of SCS and Spread Spectrum. The case of "all zero" is corresponds to uniform (homochromatic) image.

It is easy to see that the sets $\mathcal{D}_i$ are a priori interdependent via the secret key K. Indeed, the information about the secret key from these dependencies can be extracted by practical evaluation. The security level is a function of DWR, distortion compensation α, and observation index N_o. Based on the practical evaluations, we intuitively assume that security level is a logarithmic function of the observation index. Hence, the following empirical formulation is proposed for the security level of textured images.

$$h_t(K|Y^{N_o}, M^{N_o}) \cong log_2(1-\alpha)\,\Delta - 0.96 log_2 N_o + 0.35 \qquad (17)$$

To show the accuracy of the proposed empirical formulation, a large set of the well-known image databases with different sizes, resolutions, qualities and other basic characteristics are created as follow:

- USC-SIPI image database [18], composed of textures and miscellaneous images.
- TESTIMAGES archive [19].
- DIP3/e—Book Images [20], composed of images from Digital Image Processing, 3rd ed, by Gonzalez and Woods.
- CIPR Still Images [21], composed of miscellaneous images.

Security level is evaluated for these test images, embedding parameters (α, DWR) and observation index $(\mathrm{N_o})$. To show the superiority of the proposed formulation, we evaluate percentage error between the practical evaluation of security level and the previous/ proposed equations of (4)/ (17). The percentage error between practical evaluation and the previous theoretical equation $\mathrm{EP}_{\mathrm{previous,practical}}$ is defined in (18). Moreover, the percentage error between the practical evaluation and the proposed empirical equation $\mathrm{EP}_{\mathrm{proposed,practical}}$ is defined in (19). The indexes 'o', 't' and 'p' in (18) and (19), respectively, indicate the previous theoretical equation, proposed empirical equation and practical evaluation, respectively.

$$EP_{previous,practical}(\alpha.DWR, N_o) = \frac{|h_o(K|Y^{N_o}, M^{N_o}) - h_p(K|Y^{N_o}, M^{N_o})|}{h_p(K|Y^{N_o}, M^{N_o})} \qquad (18)$$

$$EP_{proposed,practical}(\alpha.DWR, N_o) = \frac{|h_t(K|Y^{N_o}, M^{N_o}) - h_p(K|Y^{N_o}, M^{N_o})|}{h_p(K|Y^{N_o}, M^{N_o})} \qquad (19)$$

Fig. 8 (a) demonstrates percentage error for $\alpha = 0.71$ as a function of DWR. Moreover, Fig. 8(b) shows the percentage error for $\mathrm{DWR} = 28\mathrm{dB}$ as a function of α. In both figures, $\mathrm{EP}_{\mathrm{previous,practical}}$ and $\mathrm{EP}_{\mathrm{proposed,practical}}$ are marked by square and circular circles, respectively. For the illustration purpose, these figures are averaged over the observation index (N_o). The simulations clearly show that the percentage error of the proposed formulation is below one percent for whole values of the embedding parameters $(\alpha.\mathrm{DWR})$. Hence, the proposed empirical equation of (17) is consistent with the practical evaluation of the entropy, while the existing theoretical analysis of (4) in the literature leads to the results far from the reality. Hence, it can be concluded that the proposed formulation is remarkably accurate for the textured images or textured regions of any image.

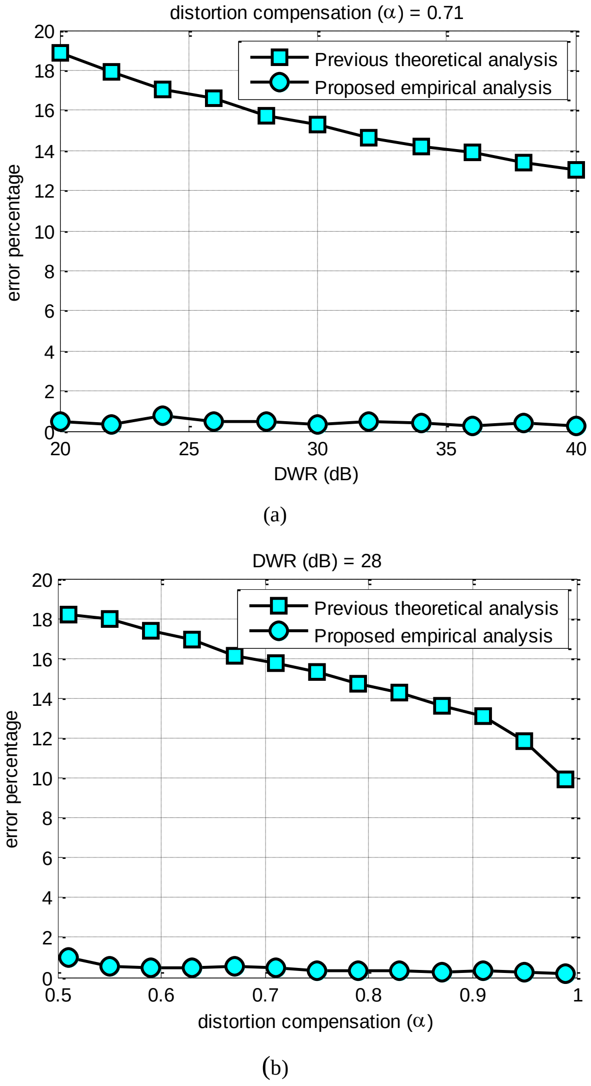


Fig. 8. Percentage error of evaluating residual entropy of SCS embedding in DCT domain in KMA scenario for test images and specified parameters; squares correspond to former theoretical equation; comparisons are done with previous theoretical results of SCS.

4.3 Security analysis for every image

Finally, the results of the uniform and textured regions are combined to propose a unified formulation for the security level of SCS in the DCT domain image watermarking. Comparison of Fig. 5 and Fig. 7 clearly shows that the security level in the DCT domain image watermarking is closely related to the uniformity of the image. Hence, we define a new variable ρ which linearly measures the uniformity of an image. The case $\rho = 1$ corresponds to the complete uniform images while $\rho = 0$ indicates almost textured images. The block diagram of computing uniformity variable ρ is illustrated in Fig. 9. It is reasonable that the unified formulation should be a combination of (11) and (17). Hence, the new theoretical-empirical formulation is states as follows:

$$\begin{aligned} h(K|Y^{N_o},M^{N_o}) &\cong \rho h_u(K|Y^{N_o},M^{N_o}) \\ &+ (1-\rho)h_t(K|Y^{N_o},M^{N_o}) \\ &= log_2(1-\alpha)\,\Delta - 1 \\ &- (1-\rho)(0.96 log_2 N_o - 1.35) \end{aligned} \qquad (20)$$

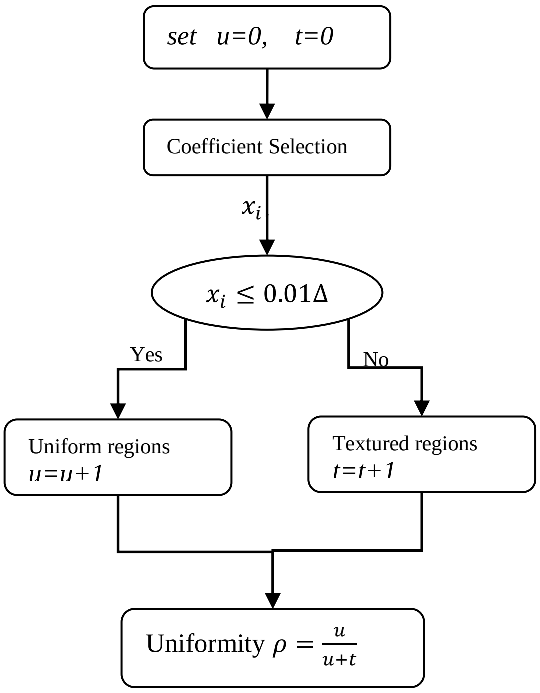
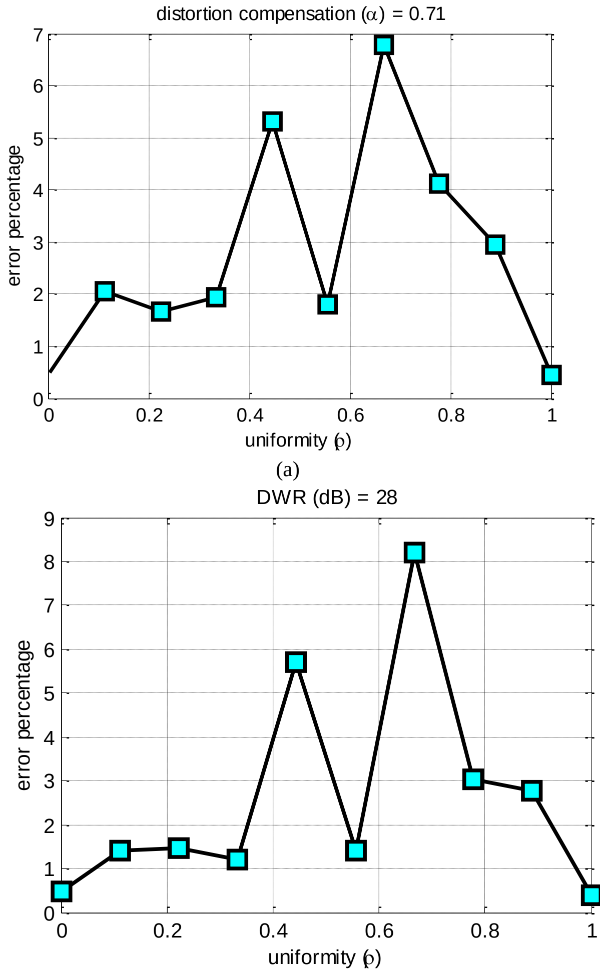


Fig. 9. Algorithm for evaluating the image uniformity.

The accuracy of the proposed analysis is verified by computing the percentage error of the proposed unified formulation and the practical security evaluation of the test images. The percentage error for several values of the embedding parameters is illustrated in Fig. 10 as a function of the uniformity ρ. Moreover, the average percentage error, computed for all values of the uniformity ρ are sketched in Fig. 11 as a function of the embedding parameters. It can be seen that the average percentage error of the proposed unified formulation is below 4% in extreme cases whereas the SCS's error is more than 27%. This figure demonstrates the accuracy and superiority of the proposed unified formulation.



(a)

(b)

Fig. 10. Percentage error of the proposed unified formulation in comparison with the practical security evaluation of the test images for some embedding parameters.

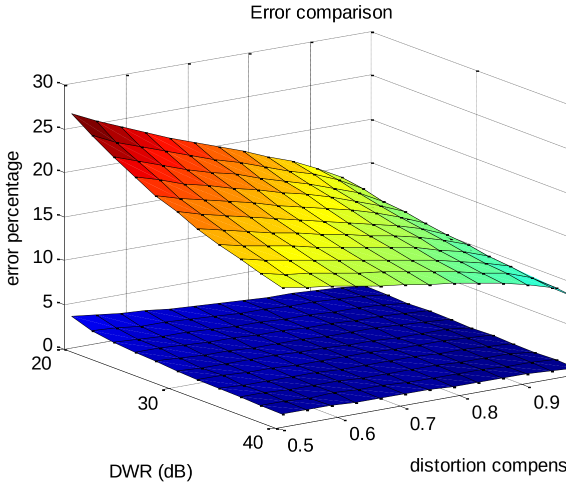


Fig. 11. Percentage error of the proposed unified formulation in comparison with the practical security evaluation of the test images for whole embedding parameters; the lighter sketch corresponds to former theoretical equation in comparison with the practical security evaluation.

5. Conclusions

This paper developed a novel information-theoretic security analysis of the scalar Costa scheme for the digital image watermarking applications. Based on the findings of this study, presence of the uniform regions in the images invalidates flat-host assumption, considered as a founded assumption by previous researches. It was shown that the flat-host assumption results in a large difference between the theoretical and practical security evaluations. To tackle this shortcoming and develop a generic analysis framework for the digital image watermarking, in this paper a theoretical equation was derived for the uniform regions while an empirical formulation was proposed for the textured (non-uniform) regions of the image. Moreover, to tackle another invalid assumption in previous works, this study assumed the observations are a priori dependent. It was demonstrated that assuming independent observation leads to the deviation of the theoretical results from the reality, even in case of non-uniform images. Finally, the theoretical equation for uniform regions and empirical equation for non-uniform regions were unified to produce a single formulation for the whole image. The final unified equation helps watermarking system designers to perform reliably designs and developments.

Acknowledgment

The authors would like to express their sincere thanks to Dr. Majedi for his valuable helps.

References

[1] M. Costa, “Writing on dirty paper (Corresp.),” *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 439- 441, 1983.

[2] J. Eggers, R. Bauml, R. Tzschoppe and B. Girod, "Scalar Costa scheme for information embedding," *IEEE Transactions on Signal Processing*, vol. 51, no. 4, pp. 1003-1019, 2003.

[3] R.M. Gray and D.L. Neuhoff, "Quantization," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2325-2383, 1998.

[4] T. Kalker, "Considerations on watermarking security," *2001 IEEE Fourth Workshop on Multimedia Signal Processing*, pp. 201-206, 2001.

[5] F. Cayre, C. Fontaine and T. Furon, “Watermarking security: theory and practice,” *IEEE Transactions on Signal Processing*, vol. 53, no. 10, pp. 3976- 3987, 2005.

[6] T. Mittelholzer, "An information-theoretic approach to steganography and watermarking," *Proceedings of the Third International Workshop in Information Hiding*, pp. 1-16, 1999.

[7] P. Bas, and T. Furon, "A New Measure of Watermarking Security: The Effective Key Length," *Information Forensics and Security, IEEE Transactions on*, vol. 8, no. 8, pp. 1306-1317, 2013.

[8] L. Perez-Freire, F. Perez-Gonzalez, T. Furon and P. Comesana, “Security of Lattice-Based Data Hiding Against the Known Message Attack,” *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 4, pp. 421-439, 2006.

[9] L. Perez-Freire, and F. Perez-Gonzalez, “Security of Lattice-Based Data Hiding against the Watermarked-Only Attack,” *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 4, pp. 593-610, 2008.

[10] B.R. Matam, and D. Lowe, "Watermark-only security attack on DM-QIM watermarking: Vulnerability to guided key guessing," *International Journal of Digital Crime and Forensics*, vol. 2, no. 2, pp. 64-87, 2010.

[11] R. Samadi, and S.A. Seyedin, "Security assessment of scalar costa scheme against known message attack in DCT-domain image watermarking," *Electrical Engineering (ICEE), 2013 21st Iranian Conference on*, pp. 1-5, 14-16 May 2013.

[12] F. Chuhong, D. Kundur and R.H. Kwong, "Analysis and design of secure watermark-based authentication systems," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 1, pp. 43- 55, 2006.

[13] J. Hernandez, M. Amado and F. Perez-Gonzalez, "DCT-domain watermarking techniques for still images: detector performance analysis and a new structure," *IEEE Transactions on Image Processing*, vol. 9, no. 1, pp. 55-68, 2000.

[14] A. Kerckhoffs, "La cryptographie militaire," *Journal des sciences militaires*, vol. 9, pp. 5-38, 1883.

[15] R. Zamir and M. Feder, “On lattice quantization noise,” *IEEE Transactions on Information Theory*, vol. 42, no. 4, pp. 1152-1159, 1996.

[16] A. Papoulis and U. Pillai, *Probability, Random Variables and Stochastic Processes*. New York: McGraw-Hill, 2002.

[17] L. Perez-Freire and F. Perez-Gonzalez, “Spread-Spectrum Watermarking Security,” *IEEE Transactions on Information Forensics and Security*, vol. 4, no. 1, pp. 2-24, 2009.

[18] USC-SIPI image database. [online]. available: http://www.sipi.usc.edu /database/.

[19] TESTIMAGES archive. [online]. available: http://www.tecnick.com/public/code/cp_dpage.php? aiocp_dp=testimages.

[20] DIP3/e—Book Images. [online]. available: http://www.imageprocessingplace.com/DIP-3E/dip3e_book_images_downloads.htm.

[21] CIPR Still Images. [online]. available: http://www.cipr.rpi.edu/ resource/stills/index.html.

Reza Samadi (M’2011) is graduated from Sharif University of technology in 2006. Now he is pursuing his PhD in Ferdowsi University of Mashhad. His research interest covers information security, physical layer security, and applications of information theory.

Seyed Alireza Seyedin was born in Mashhad. He received the B.S. degree in Electronics Engineering from Isfehan University of Technology, Isfehan, Iran in 1986, and the M.E. degree in Control and Guidance Engineering from Roorkee University, Roorkee, India in 1992, and the Ph.D. degree from the University of New South Wales, Sydney, Australia in 1996. He has been an Associate Professor with the Department of Electrical Engineering, the Ferdowsi University of Mashhad, Mashhad, Iran. His research interest includes image processing, computer vision, signal processing, and pattern recognition. In these fields, specially, he is interested in image analysis, motion detection and estimation in image sequences, autonomous vehicles, and diverse applications of the radon transform.