Channel Capacity Analysis of Spread Spectrum Audio Watermarking

M. Khademi

Abstract—**In this paper, information-theoretic analysis of spread spectrum watermarking (SSW) is performed. The analysis is focused on deriving channel capacity of SSW system in order to determine the most embedding rate with the highest robustness. Channel capacity is calculated for three cases when the watermark channel introduces no attack, the additive noise attack and the de-synchronization attack. The analyses in this paper are based on the practical aspects and the embedding and detecting strategy of SSW. In each section, first the channel capacity of SSW is calculated and then all the effective parameters on channel capacity are examined.**

Index Terms—**Channel capacity analysis, spread spectrum watermarking (SSW), interference noise of host, water-filling approach, de-synchronization attack.**

I. INTRODUCTION

DURING the last decade, digital watermarking has received considerable attention from security and digital signal processing groups. Digital watermarking is the process of embedding bits of information into a multimedia signal, which is called the host signal, in a way that no perceivable distortion is produced. Many watermarking schemes have been proposed. However, each scheme is appropriate for a particular application. To compare watermarking schemes to choose the appropriate one according to the restriction of applications, some watermarking features have been proposed. Theses features are: perceptual transparency, security, embedding rate and robustness. By embedding rate we mean the maximum bit rate of information which can be embedded into a host signal without any perceivable distortion. The robustness is also the ability of watermarking scheme for the reliable detection of embedded information after performing attack manipulations through a watermark channel [1]. In the watermarking literature, the attack is referred to as manipulations that would like to remove any trace of the embedded information from the watermarked signal without introducing perceptual distortions. Hence, attack could be unintentional (e.g. additive noise, compressions, A/D and D/A) and intentional (e.g. De-synchronization and active steganalysis methods).

Currently, enhancing the robustness and the maximum embedding rate is one of the most desired aspects of each watermarking scheme. However, analysis shows that the embedding rate is in complete conflict with the robustness. The important role of theses features and also the conflict between them are the main motivation for emerging the application of information theory in watermarking. The aim of information-theoretic analysis of watermarking and particularly calculating channel capacity of watermarking schemes is to compute theoretically the maximum number of embedded information bits that can be reliably embedded and detected from the host signal.

In addition to attack manipulations the host interference noise has the same effect as attack for the false detection of embedded information. The host interference noise is actually the effect of the host signal on hindering the correct detection of embedded information. Based on the host interference effect, watermarking schemes can be categorized into two classes. Schemes of the first class do not utilize any knowledge about the host signal but perform embedding based on its statistics to satisfy the necessary embedding distortion constraint. These methods are called host interference non-rejecting methods (like Spread spectrum watermarking and echo hiding for audio watermarking algorithms). Watermarking schemes of the second class, regarded to as host interference rejecting methods, are based mostly on the quantization operation (like QIM, Patchwork and LSB coding for audio watermarking algorithms). It was shown that host interference rejecting schemes have significantly higher embedding rate than host interference non-rejecting schemes due to the host interference cancellation [15]. However, the schemes of the later class have a high degree of robustness against attacks in comparison with the second methods [5].

In this paper, an information-theoretic analysis of the spread spectrum watermarking (SSW) scheme is presented. SSW is the most promising watermarking scheme that offers a high degree of robustness and relatively low embedding rate for the embedding of the bits of information. The SSW scheme also belongs to the class of host interference non-rejecting methods. Hence, in addition to external attack introduced by the watermark channel, the host interference noise should be considered for the channel capacity analysis of SSW.

Costa's paper entitled as "Writing on Dirty Paper" introduced the channel capacity analysis of watermarking systems [2]. Costa considered a communication problem in which a message is sent from an encoder to a receiver through an interference noise which is generated at the encoder. The channel capacity for the case the interference noise is known to the encoder but not known at the receiver is analyzed. The analysis is regarded as the basic theorem in information theoretic analysis of watermarking. In [3] a survey of information theoretic analysis of watermarking is performed which classifies proposed schemes are classified into three groups: algorithm analysis, game approach, and attack modeling.

Manuscript received October 13, 2009; revised July 3, 2010.

The author is with the Electrical Engineering Department, Ferdowsi University of Mashhad, Iran (e-mails: khademi@um.ac.ir).

Publisher Item Identifier S 1682-0053(10)1842

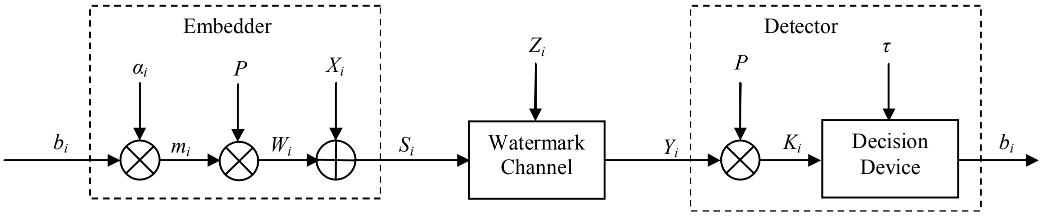

Fig. 1. Transmission and detection of each bit of information in the SSW scheme.

The class of algorithm analysis comprises papers that compute performance bounds transmission rate of embedded information for a specified watermarking scheme. Voloshynovskiy *et al.* performed channel capacity analysis of watermarking for real images (where the samples are not considered i.i.d.), and the schemes with different strategy of side information watermarking including distortion compensated dither modulation and SSW for image watermarking [4] and [5]. Barni et al performed the analysis for DCT-based and DFT-based image spread spectrum watermarking against the interference noise of host [6] and [7].

The game analysis performs an analysis of game for a watermarking scheme. The theory of the game considers the watermarking problem in which two parties are involved: an embedder and an attacker. While the embedder tries to maximize the embedding rate, the attacker attempts to minimize it. Moulin et al explored a perfect analysis of watermarking capacity game in [8], [9] and [10]. The paper computes the watermarking game capacity for different attack scenarios and also different distortion constraints by developing Costa's results. Furthermore, the best scenarios for attack and watermark embedding were exploited. O' Sullivan et al investigated the capacity of the game for private and public watermarking strategies [11]. Cohen and Lapidoth [12] derived a capacity formula of the watermarking game for Gaussian covertext and an almost-sure squared-error distortion constraint.

Performing attack modeling analysis is restricted to a particular attack. In these papers a specified attack is modeled and its effect on reliable transmission rate over the different watermarking systems has been investigated. Zaidi et al examined audio watermarking under the de-synchronization and additive noise attacks [13]. In [20] the specific problem of channel capacity analysis in the face of quantization-based compression and the best watermarking strategy against the attack were considered by Kundur. The analysis was developed to different compressions by Fei et al in [14]. They also considered different algorithms of image watermarking for JPEG compression attack.

The approaches in this paper belong to the first group of the three categorized previous works. In this paper, the computations on the channel capacity are based on the embedding and the detecting strategy of the SSW scheme. Therefore, in addition to computing the channel capacity in different situations, the effect of all the parameters on embedding and detection of information is examined.

The rest of the paper is organized as follows. In Section II a description of SSW is performed in order to introduce the exact SSW approach and also the notations in the rest of the paper. Derivation of formulae of the channel capacity of SSW without any attack is described in Section III. Section IV describes channel capacity under the noise attack, and Section V describes channel capacity under the de-synchronization attack. Conclusions are dealt with in Section VI.

II. REVIEW OF SPREAD SPECTRUM AUDIO WATERMARKING

In this paper we use the following notation. Vectors are denoted by bold letters (e.g. $\mathbf{x}$), their elements by the italic letters (e.g. x), and random variables are denoted by capital letters (e.g. X). Cox *et al* popularized SSW in [15] and then developed the technique as communication with side information [16]. In [16] also a survey on the SSW proposed schemes has been performed. To clarify the exact SSW approach used for our analysis, a brief description of the SSW scheme is presented in this section.

A. The Structure

The structure of the SSW scheme is illustrated in Fig. 1. The embedding strategy is base on the spread spectrum techniques. Assume that an embedder is about to embed a collection of *n* bits of information $M = \{b_1 ,.., b_n\}$ into a host signal $\mathbf{x}$. Here b_i denotes the *i*th bit of the embedded information. Let $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, ...)$, where each vector $\mathbf{x}_i = (x_i(1) ,..., x_i(N))$ contains N samples of the host signal. The embedder embeds each bit $b_i \in \{-1,1\}$, $1 \le i \le n$ into the *i*th vector of the host signal $\mathbf{x}_i$, to produce the *i*th vector of the watermarked signal $\mathbf{s_i}$, following the steps below

1. Attenuate the power of b_i by a value $\alpha_i \in [0,1]$ to reduce the possibility of any perceivable distortions after the embedding procedures.

   $$\mathrm{m_i} = a_i b_i \ , \ \ i = 1, 2, ...n \tag{1}$$

2. Generate one period of pseudo-random noise sequence $\mathbf{p} = [p(1),..., p(N)]$. Here, $p(l)$, $1 \le l \le N$, denotes the l-th chip of $\mathbf{P}$.
3. Produce the *i*th vector of the watermark signal

   $$\mathbf{w}\mathrm{i} = \mathrm{m_i} \times \mathbf{p} \tag{2}$$

4. Add the watermark vector $\mathbf{w}_\mathrm{i}$ to the host vector $\mathbf{x}_i$ to obtain the watermarked vector $\mathbf{s}_\mathrm{i}$

   $$\mathbf{s}_i = \mathbf{w}_i + \mathbf{x}_i \tag{3}$$

After the watermarked signal $\mathbf{s}_\mathrm{i}$ is generated, it is transmitted to the watermark detector through a watermark channel. Since the power of the message is spread through the whole frequency bandwidth of the host signal,
the embedding procedures do not produce any perceivable distortion.

The watermark channel is subject to attacks that attempt to remove any trace of m_i from $\mathbf{s}_\mathrm{i}$. The attack is actually a signal z that is added by an attacker. Therefore, the received signal to the detector $\mathbf{y}_\mathrm{i}$ is $\mathbf{y}_i = \mathbf{w}_i + \mathbf{x}_i + \mathbf{z}_i$. where $\mathbf{z}_\mathrm{i}$ is the i-th vector of $\mathbf{z}$ containing N samples.

At the receiving side, the detector extracts the message bit b_i from $\mathbf{y}_\mathrm{i}$, using the correlation property of PN sequences. Let $\mathbf{p}_\mathrm{i}$ and $\mathbf{p}_\mathrm{j}$ be two maximal length PN sequences with the length N. Then the cross correlation between them is

$$Cr(\mathbf{P}_i,\mathbf{P}_j)=\frac{1}{N}\mathbf{p}_i\mathbf{p}_j^{\,T} = \frac{1}{N}\sum_{l=1}^{N} p_i(l)p_j(l) = \begin{cases} 1 & \text{if } i=j \\ -\dfrac{1}{N} & \text{Otherwise} \end{cases}$$

Here, $p_i(l)$ and $p_j(l)$, $1 \le l \le N$ denote each chip of $\mathbf{P}_\mathrm{i}$ and $\mathbf{P}_\mathrm{j}$ and $Cr(.,.)$ is a correlation function between two vectors. We now compute the cross correlation between $\mathbf{w}_\mathrm{i}$ and the $\mathbf{P}$ used at the embedder

$$Cr(\mathbf{w}_i,\mathbf{p}) = \frac{1}{N}\mathbf{w}_i\mathbf{p}^T = \frac{1}{N}\sum_{l=1}^{N} m_i p(l)p(l) = m_i \tag{4}$$

Therefore the cross correlation between the received signal to the detector $\mathbf{y}_\mathrm{i}$ and the PN sequence $\mathbf{p}$ used at the embedder is

$$\begin{aligned} Cr(\mathbf{y}_i,\mathbf{p}) &= \frac{1}{N}\mathbf{y}_i\mathbf{p}^T = \frac{1}{N}\left(\mathbf{w}_i + \mathbf{x}_i + \mathbf{z}_i\right)\mathbf{p}^T = \\ \alpha_i m_i &+ \frac{1}{N}\mathbf{x}_i\mathbf{p}^T + \frac{1}{N}\mathbf{z}_i\mathbf{p}^T = \alpha_i m_i + K_{ix} + K_{iz} \end{aligned} \tag{5}$$

where K_{ix} and K_{iz} are the i-th value of K_x and K_z as follows

$$K_{ix} = \frac{1}{N}\sum_{l=1}^{N} x_i(l)p(l),\; K_{iz} = \frac{1}{N}\sum_{l=1}^{N} z_i(l)p(l) \tag{6}$$

By comparison with a threshold (τ), the b_i is extracted by a decision device. Since $m_i \in \{-\alpha_i, \alpha_i\}$ and K_{ix} and K_{iz} are zero mean signals because of being zero mean X, τ is considered zero. It should be noted that in [23] an intelligent algorithm for determining τ has been proposed which specifies τ dynamically. The parameter τ is calculated according to the dynamic range of X.

As (5) shows, signals K_x and K_z are the interference signals for the correct detection of m. Even in the case $Z=0$, K_x will remain and it would disturb the detection. Since K_x is only dependant on X it is called the *host interference* in this paper.

B. Channel Capacity

As mentioned before, "*Writing on dirty paper*" [2] is the premium work on information theoretic analysis of watermarking. It is proved by Costa that the channel capacity for transmission of hidden information with the code word W through a watermarking system is obtained as follows

$$C = \max\; [I(U;Y \mid H) - I(\,U;X \mid H)] \tag{7}$$

where the maximum is over all joint distributions of the form $P_r(x,\, h)P_r(u,\, w \mid x,\, h)P_r(y \mid w,\, x,\, h)$ and the lowercase letters denoting the individual values of their capital letter random variables and P_r denoting probability mass function. In (7), $I(.)$ stands for the mutual information function, $(\mathrm{XUH}) \rightarrow \mathrm{S} \rightarrow \mathrm{Y}$ form a Markov chain and U is the auxiliary variable for embedding hidden information as follows.

$$U = W + \eta X \tag{8}$$

where η is the parameter that adopts embedding according to X. The (7) is derived as a general formula and is applicable for all the watermarking algorithms.

SSW does not use any information about X, η=0. Hence,

$$C = \max\; [I(W\;;\;Y \mid P) - I(\;W;X \mid P)] \tag{9}$$

Extending (9) using standard information theoretic rules $C = \max[H(Y \mid P) - H(Y \mid W,P) - H(X \mid P) + H(X \mid W,P)]$ Since X is independent of W and P

$$H(X \mid W,P) = H(X) \tag{10}$$

Let CR denote the random value of the cross correlation between Y and P. Then

$$H(Y \mid P) = H(CR) \tag{11}$$

Since, $\mathbf{wi} = m_i\mathbf{P}$

$$H(Y \mid W,P) = H(CR \mid M) \tag{12}$$

Therefore, the channel capacity of SSW is evaluated as follows

$$C = \max\;\; [H(CR) - H(CR \mid M)] \tag{13}$$

where the maximum is taken over all the distributions of $P_r(x,w)P_r(y \mid w,x)$.

We now explain the constraint in (13). As mentioned, the distance between the power of $\mathbf{w}_\mathrm{i}$ and $\mathbf{x}_\mathrm{i}$ should be preserved by the parameter α_i to keep the distortions below the human auditory system (HAS). The distance is denoted by the parameter *HWR* (Host to Watermark ratio), which is the power of host to the power of watermark ratio as follows

$$HWR = \frac{Pow(\mathbf{x}_i)}{Pow(\mathbf{w}_i)} = \frac{Pow(\mathbf{x}_i)}{\alpha_i^{\,2}} \tag{14}$$

where $Pow(.)$ stands for the power of the signal. The HWR is assumed to be constant in S and equal to a predefined value. Therefore, the constant value of HWR is the constraint for performing the channel capacity calculations.

Comparing (13) with the formula for channel capacity of a binary symmetric channel (BSC) would be very useful. The channel capacity of a BSC with B_O as the output signal and B_I as the input signal is as follows

$$C = \max\left[H(B_O) - H(B_O \mid B_I)\right] \tag{15}$$

Comparing (13) and (15) shows that the behavior of the whole blocks of the embedder and the detector of SSW is similar to that of a BSC, which accepts m as the input signal, adds K_x and K_z as the noise signal to give the sum as the output signal (K). Hence, SSW system could be modeled as a BSC which is depicted Fig. 2. This modeling would simplify our calculations for obtaining channel capacity.

III. CHANNELS WITHOUT ATTACK

The purpose of this section is to obtain maximum reliable transmission rate of embedded information using

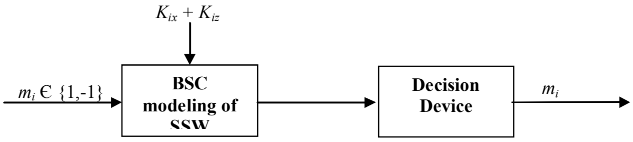

Fig. 2. The BSC modeling of the SSW scheme.

the SSW scheme in case where the watermarked signal passes through a transparent channel. A transparent channel is one that does not impose any attack on the transmission of the watermarked signal. Obviously, in this case the only effective factor on the transmission of the embedded information is the host interference noise K_x.

A. Channel Capacity

In the case where transmitting S from a transparent channel occurs $K_z = 0$. To compute the channel capacity using (13), the cross correlation Cr at the detector must be computed. Using (5), the cross correlation $Cr(Y_i, P)$ is

$$Cr(Y_i, P) = M_i + K_{ix} \tag{16}$$

Hence, the channel capacity is calculated as follows

$$C = \max[H(M + K_x) - H(K_x)] \tag{17}$$

Also, according to the BSC modeling of SSW, this situation behaves as a BSC that adds K_x as the noise signal to M. Using (13) for calculating channel capacity and also the BSC modeling of SSW determines the application of Shannon's channel capacity formula for deriving the capacity. Employing the Shannon's formula would be as follows

$$C = B\log_2\left[1 + \frac{Pow(M)}{Pow(K_x)B}\right] \text{ bit/ sec} \tag{18}$$

where B is the bandwidth of transmitted M. Since $M_i \in \{-\alpha_i, \alpha_i\}$, the following is obtained

$$Pow(M_i) = \alpha_i^2 \tag{19}$$

As (18) shows, statistic characteristics of K_x should be computed for deriving the formula of the channel capacity. Since X is an audio signal, it can be considered to be a zero mean signal. Therefore, K_x is zero mean. Autocorrelation of K_x is as follows

$$R_i(j) = E\left[K_{ix}K_{(i+j)x}\right] = \frac{1}{N^2}\left\{\sum_{k=1}^{N} E\left[x_i(k)x_{i+j}(k)\right]\right\} \quad j = 1,2,... \tag{20}$$

Although there is a high correlation between the adjacent samples of an audio signal, the expectation in (20) is performed between the samples that have distances equal to multiple value of N. Since N is usually chosen very large, there is a very low level of correlation between the samples with such distances and therefore K_x could be considered as a white Gaussian noise. Hence, evaluating $R_i(j)$ shows that without using the wrong assumption that samples of host are i.i.d. random variables, K_x samples remain i.i.d.. Using (20) $\mathrm{Pow}(K_{ix})$ is calculated as follows

$$Pow(K_{ix}) = R_i(0) = \frac{1}{N^2}E\left\{\sum_{k=1}^{N}\left[x_i(k)^2\right]\right\} = \frac{Pow(X_i)}{N} \tag{21}$$

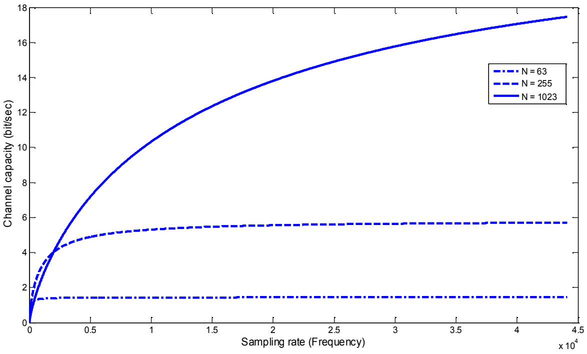

Fig. 3. The effect of increasing the sampling rate on the channel capacity of SSW.

Hence,

$$\frac{Pow(M_i)}{Pow(K_{ix})} = \frac{N}{HWR} \tag{22}$$

and for all i, (21) would be constant in all regions of S. Since each period of P or N samples of S_i contains one bit of embedded information, the maximum rate (r) for embedding the bits of information into the host equals to what follows

$$r = \frac{f_x}{N} \tag{23}$$

where f_x is the sampling rate of the audio host signal and the bandwidth of transmitted M would be as follows

$$B = \frac{r}{2} = \frac{f_x}{2N} \tag{24}$$

By replacing the parameters obtained above, the channel capacity of SSW under the situation of the transparent watermark channel is obtained in the follow way

$$C = \frac{f_x}{2N}\log_2\left[1 + \frac{2N^2}{HWR.f_x}\right] \; bit/\sec \tag{25}$$

B. Simulation Results

Eq. (25) shows that f_x, N and HWR are the effective parameters on computing the channel capacity of SSW under the transparent channel. In this section the effect of these parameters on C are examined.

We first examine the effect of the sampling rate of the host signal on the channel capacity. Fig. 3 illustrates the Effect of increasing f_x on the channel capacity C. In this Figure, C is calculated for the case the values of N are 63, 255 and 1023. The parameter HWR is set to 18 dB to guarantee unperceivable distortion of embedding the bits of information. As the Figure depicts, increase in f_x enhances the channel capacity C. However, the effect of increasing f_x on the channel capacity degrades with lower values of N such that increase in f_x loses its effect on C when $N < 63$. Therefore, SSW with the music signal which has the standard sampling rate of 44100 sym/ sec can embed larger bits of information than the SSW with the speech signal which has the sampling rate of 8000 sym/ sec.

The other effective parameter of SSW on the channel capacity is the length N of the PN sequence. Considering the embedding procedures of SSW, increase in N raises

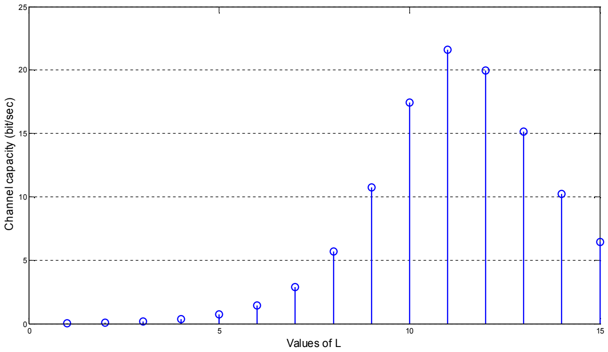

Fig. 4. The effect of increasing the length of PN sequence on the channel capacity of SSW.

the robustness of SSW against K_x. However, enhancing N decreases the maximum number of bits of information that can be embedded into the hoist signal. Hence, there is a trade-off between the robustness and the embedding rate when N increases. We anticipate a peak for the channel capacity of SSW versus different values of N. Since P should be a maximal length PN-sequence for employing in SSW, N is computed according to the following formula.

$$N = 2^{L} - 1 \qquad L = 1, 2, \ldots \tag{26}$$

Fig. 4 illustrates the channel capacity of SSW for different values of N. As the Figure depicts, the peak occurs at $N = 2047$. Hence, a PN-sequence with the length $N = 2047$ is the optimal to use in SSW. f_x and HWR used in Fig. 4 were set to $f_x = 44100$ sym/sec and $HWR = 18\,\text{dB}$.

We now examine the effect of increasing HWR on the channel capacity of SS. The parameter HWR should be set higher for the music signal in comparison with the speech signal. The reason is that the quality of music must be excellent due to the very low level expectation of noise by the listener. However, speech signals mostly passes through a noisy environment and a level of noise for speech is always expected. This makes the lower quality of the speech signal rational.

Fig. 5 illustrates the channel capacity of SSW versus different values of HWR for both speech and music signals. The range of HWR considered for speech signal is between 30 (14.7 dB) to 60 (17.7 dB). The reason is that HWR above 30 has the mean opinion score (MOS) of above 4 [18] which is expected as a good quality for speech. However, HWR used for music should be in higher values and the range between 60 (17.7 dB) to 90 (19.5 dB) to guarantee the MOS of above 5. Although higher HWR used for music decreases the capacity, since f_x used for music is higher than speech it keeps the channel capacity of music in a higher position. The value of N used in this Figure is 2047 chips.

IV. CHANNEL CAPACITY ANALYSIS UNDER THE ADDITIVE NOISE ATTACK

In this section the effect of additive Gaussian noise in the performance of SSW channel capacity is analyzed. Additive noise is the most well-known attack in the watermark channel as an unintentional attack to impair the watermark signal.

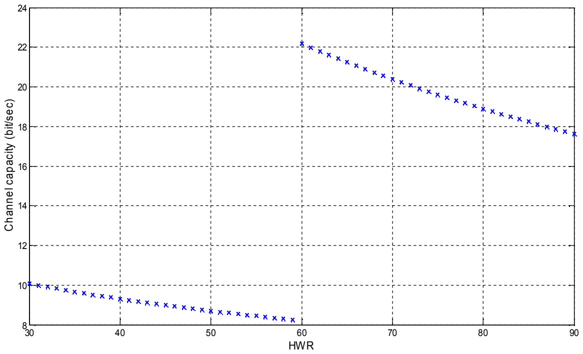

Fig. 5. The effect of increasing HWR rate on the channel capacity of SSW.

A. Channel Capacity

Employing the procedures of deriving the formula for the channel capacity of SSW according to the preceding section, in the case of additive noise attack, K should be calculated. Under the additive noise attack K appears to be exactly the same as (5) and K_z has been resulted from the additive noise (Z) as follows

$$Z \sim N \quad (0, \; V) \tag{27}$$

Using (13), the channel capacity could be derived as follows

$$C = \max \; [H(M + K_x + K_z) - H(K_x + K_z)] \tag{28}$$

The BSC modeling similar to Fig. 2 can also be used. Using the model, SSW accepts M, adds K_x + K_z as the noise signal to give the sum as the output signal. Using the same approaches in the transparent channel for calculating the channel capacity, the auto correlation of the noise signals on detecting *m* as K_x + K_z should be evaluated as follows

$$R_i(j) = E\left[K_{ix}K_{(i+j)x} + K_{iz}K_{(i+j)z}\right] = \frac{1}{N^2}\left\{\sum_{k=1}^{N} E\left[x_i(k)x_{i+j}(k)\right] + \sum_{k=1}^{N} E\left[z_i(k)z_{i+j}(k)\right]\right\} \quad j = 1, 2, \ldots \tag{29}$$

Employing the same analysis in the preceding section and considering the fact that Z is a white Gaussian noise, the samples of the autocorrelation remains i.i.d.. Hence, $Pow(K_{ix})$ is calculated as follows

$$Pow(K_{ix} + K_{iz}) = R_i(0) = \frac{1}{N^2} E\left\{\sum_{k=1}^{N}\left[x_i(k)^2\right]\right\} + V = \frac{Pow(X_i) + V}{N} \tag{30}$$

Hence

$$\frac{Pow(M_i)}{Pow(K_{ix} + K_{iz})} = \frac{N\alpha_i}{Pow(X_i) + V} = \frac{N}{HWR + \dfrac{V}{\alpha_i}} \tag{31}$$

Eq. (31) shows that in contrast with (22), the power of input signal to the noise signal changes with time. The reason is that the power of Z is constant and does not change with to the power of M. Hence, Shannon's channel capacity formula for cannot not be applied to compute the channel capacity of the SSW scheme is noisy environments. Instead, the water-filling analysis which is useful for the channels with colored noise should be employed here [19].

TABLE I
THE CHANNEL CAPACITY OF THE SSW SCHEME WITH A *MUSIC* HOST SIGNAL UNDER THE ADDITIVE WHITE GAUSSIAN NOISE ATTACK

N WNR	31	63	127	255	511	1023	2047	4095	8191
20 db	0.59	1.2	2.4	4.6	9.12	15.2	19.6	18.7	14.4
10 db	0.59	1.2	2.4	4.6	9.12	15.2	19.6	18.7	14.4
0 db	0.58	1.29	2.39	4.5	9	15.1	19.4	18.6	14.4
-10 db	0.51	1.1	2.1	4.3	8.2	13.8	18.2	17.7	14
-20 db	0.26	0.57	1.1	2.1	4	7.5	11.4	13	11.3

TABLE II
THE CHANNEL CAPACITY OF THE SSW SCHEME WITH A *SPEECH* HOST SIGNAL UNDER THE ADDITIVE WHITE GAUSSIAN NOISE ATTACK

N WNR	31	63	127	255	511	1023	2047	4095	8191
20 db	1	2	3.8	7	10.1	10.8	9	6.9	4.2
10 db	1	2	3.8	7	10.1	10.8	9	6.9	4.2
0 db	1	1. 9	3.8	6.8	10	10.7	8.9	6.4	4.1
-10 db	0.8	1.6	3.3	5.8	8.8	9.8	8.5	6.2	4
-20 db	0.03	0.62	1.1	2.4	4.2	5.7	5.9	4.7	3.3

Using the water filling approach, calculating the channel capacity of SSW under the additive noise attack (Z) is performed. By using the water filling analysis, channel capacity calculation could be performed as follow [19]

$$C=\int_0^B \frac{1}{2}\log_2\left[1+\frac{(J-p_z(f))^+}{P_z(f)}\right] \tag{32}$$

where B denotes bandwidth of transmitted M, $P_z(f)$ is the power spectral of (K_x+K_z) according (33), $(\ .\ )^+$ denotes the positive part as (34) and J is a constant that is chosen in order to satisfy (35)

$$P_z=PSD(K_x+K_z) \tag{33}$$

$$(L)^+=\begin{Bmatrix} L & if \quad L>0 \\ 0 & if \quad L<0 \end{Bmatrix} \tag{34}$$

$$\int_0^B (J-P_z(f))^+ df = Pow(M) \tag{35}$$

B. Simulation Results

It was discussed that using (32) the channel capacity of SSW under the additive noise attack could be calculated. However, since as (32) shows, the channel capacity is completely dependant on the frequency spectrum of X. Therefore, a general conclusion considering the entire host signals could not be derived. In this Section, using the computer simulations the channel capacity calculations for different situations is performed in order to give some values of the channel capacity of SSW in bit per second. Since channel capacity completely depends on spectral frequency features of X the simulations is performed for a typical music and a typical speech signal.

Table I illustrates the channel capacity of SSW which is derived by simulations and using a typical music signal as the host signal. The simulations have also been derived for different values of N and WNR, which is the watermark to noise ration as (36), to give a better insight for the comparison.

$$WNR=\frac{Pow(W)}{Pow(Z)}=\frac{Pow(W)}{V} \tag{36}$$

where $Pow(W)$ is the average power of all W_i. Also, to obtain the simulations results in Table I f_x is set to 44100 sym/sec (since the signal is music) and HWR used for this table is set to 75 (18.75 dB). Table I shows that for the range of $WNR>0$ dB, SSW is very robust against the additive noise attack. However, for the range of $WNR<0$ dB, the robustness fails drastically. It also specifies that as WNR decreases, the amount of N that produces the peak of channel capacity moves to a higher value gradually, so that for $WNR<-20$ dB, the peak occurs at $N=4095$. The reason is that for lower WNR, the robustness would be more important in comparison with the embedding rate for increasing channel capacity.

Table II illustrates the channel capacity which is derived using the SSW parameters similar to Table I but for speech signals. In this Table also the range of WNR and N is the same as Table I. As the Table shows, according to the previous results, the channel capacity of music SSW is much higher than speech SSW and also the noise resistance of SSW system would increase by enhancing WNR. f_x is set to 8000 sym/sec and the HWR used for this Table is set to 45. As the Table indicates, all the conclusions regarding Table I could be exploited from Table II.

The comparison of the two Tables shows that the dynamic rage of the variation of channel capacity by increasing N is much higher for the music signal. Also for the amounts of N below 1023, the channel capacity of speech signals would be higher than music signal. The reason is that the HWR for speech signals is much less than the HWR used for music signals.

In APPENDIX a more general analysis is performed using the approximations and a special case.

V. CHANNEL CAPACITY UNDER THE DE-SYNCHRONIZATION ATTACK

In this section the channel capacity analysis of SSW under the de-synchronization attack is performed. While SSW is very resilient against some attacks such as scaling and a range of additive noise, it is very sensitive against this attack. The de-synchronization attack is performed using a delay in the samples of S. The delay causes the detector not to detect hidden information precisely because two PN-sequences for correlation function should be synchronized exactly and the delay would damage the correlation and the receiver may not detect correctly.

A. Channel Capacity

This section is to obtain the channel capacity of SSW under the de-synchronization attack. The de-synchronization attack is not as simple as the additive noise and a channel modeling is essential. Hence, before investigating the channel capacity analysis of SSW regarding the de-synchronization attack, an appropriate modeling should be considered in order to demonstrate the attack according to the added signal to S, which is denoted by Z in (5). In [22] the channel modeling of de-synchronization attack based on ISI approach has been studied. In this reference it is proved that using the ISI modeling a delay $\tau = \delta T$ as the result of de-synchronization attack changes S according as follows

$$Y(n) = S_j(n) = \sin c(\delta)S(n) + \sum_{k=-\infty}^{+\infty} S(k)\sin c(n-k+\delta) \quad (37)$$

where the index j denotes the jittered signal. Employing the modeling, the received signal to the detector would be the attenuated watermarked signal by $\mathrm{sinc}(\delta)$ and the added attack signal (Z) as follows

$$Z(n) = \sum_{k=-\infty}^{+\infty} S(k)\mathrm{sinc}(n-k+\delta) \quad (38)$$

Therefore, using the ISI channel modeling the de-synchronization attack could be presented by the additive signal Z as (38). To calculate the channel capacity using the approaches like the previous analysis K_i should be computed. In this case K_i would be as follows

$$\begin{aligned} K_i &= \frac{1}{N}Y_i P = \frac{1}{N}S_{Ji}P = \frac{1}{N}\left(W_{ji}P + X_{ji}P\right) = \\ &\frac{1}{N}\mathrm{sinc}(\delta)\alpha_i \sum_{k=1}^{N} w_i(k)P(k) + \\ &\frac{1}{N}\sum_{k=1}^{N}\left[\sum_{l=-\infty}^{+\infty} \alpha_i w_i(l)\mathrm{sinc}(k-l+\delta)\right]P(k) + \\ &\frac{1}{N}\mathrm{sinc}(\delta)\sum_{k=1}^{N} x_i(k)P(k) + \\ &\frac{1}{N}\sum_{k=1}^{N}\left[\sum_{l=-\infty}^{+\infty} x_i(l)\mathrm{sinc}(k-l+\delta)\right]P(k) \end{aligned} \quad (39)$$

Extending (39) would be as follows

$$\begin{aligned} K_i = &\frac{1}{N}[\mathrm{sinc}(\delta) + \mathrm{sinc}(\delta+1) + \ldots + \\ &\mathrm{sinc}(\delta+N)]\alpha_i M_i + \mathrm{sinc}(\delta)K_{ix} + K_{iz} \end{aligned} \quad (40)$$

Hence, the autocorrelation of $K_{ix} + K_{iz}$ is as follows

$$\begin{aligned} R_i(j) = &\mathrm{sinc}^2(\delta)E\left[K_{ix}K_{(i+j)x}\right] + \mathrm{sinc}(\delta)E\left[K_{ix}K_{(i+j)z}\right] \\ &+ \mathrm{sinc}(\delta)E\left[K_{(i+j)x}K_{iz}\right] + E\left[K_{iz}K_{(i+j)z}\right] \end{aligned} \quad (41)$$

The fist term is equal to (20) and the others are as follows

$$\begin{aligned} &E\left[K_{ix}K_{(i+j)z}\right] = \\ &\frac{1}{N^2}E\left\{\sum_{k=1}^{N} x_i(k)\left[\sum_{l=-\infty}^{\infty} x_{i+j}(l)\mathrm{sinc}(k-l+\delta)\right]\right\} \end{aligned} \quad (42)$$

$$\begin{aligned} &E\left[K_{(i+j)x}K_{iz}\right] = \\ &\frac{1}{N^2}E\left\{\sum_{k=1}^{N} x_{i+j}(k)\left[\sum_{l=-\infty}^{\infty} x_i(l)\mathrm{sinc}(k-l+\delta)\right]\right\} \end{aligned} \quad (43)$$

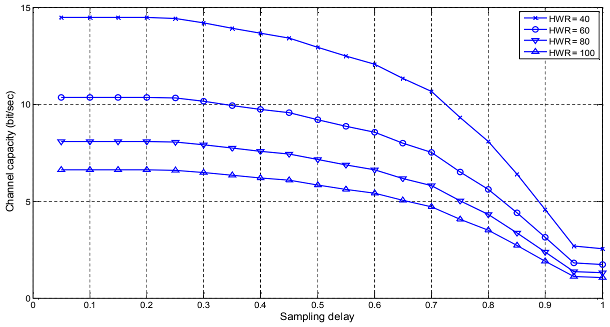

Fig. 6. The effect of increasing the sampling delay in the de-synchronization attack on the channel capacity of SSW.

$$\begin{aligned} E\left[K_{ix}K_{(i+j)z}\right] = &\frac{1}{N^2}E\{\sum_{k=1}^{N}\left[\sum_{l=-\infty}^{\infty} x_i(l)\mathrm{sinc}(k-l+\delta)\right] \\ &\left[\sum_{n=-\infty}^{\infty} x_{i+j}(n)\mathrm{sinc}(k-n+\delta)\right]\} \end{aligned} \quad (44)$$

Since, $\sin c(d) \le 0.01,\ \forall d,\ d \ge 30$ a very low correlation between the samples of $R_i(j)$ is expected. Using 40 $Pow(K_{ix} + K_{iz})$ is calculated as follows

$$\begin{aligned} &Pow(K_{ix} + K_{iz}) = R_i(0) = \\ &\frac{1}{N}\left[\mathrm{sinc}^2(\delta) + (1 - \mathrm{sinc}^2(\delta))\right]Pow(X_i) = \frac{1}{N}Pow(X_i) \end{aligned} \quad (45)$$

The reason is obvious by referring to the fact that a delay in the samples of a signal does not change the energy of the signal. Examining (39) and (44) shows that the approaches for calculating the channel capacity of SSW in the case of additive noise attack are very similar to the approaches in the case of transparent channel. Hence, by using the same analysis used in Section III, the channel capacity under the de-synchronization attack can be calculated as follows

$$C = \frac{f_x}{2N}\log_2\left[1 + \frac{2N^2\mu}{(HWR)f_x}\right] \quad (46)$$

where μ is as follows

$$\mu = \frac{1}{N}[\mathrm{sinc}(\delta) + \frac{1}{N}\mathrm{sinc}(\delta+1) + \frac{1}{N}\mathrm{sinc}(\delta+2) + \ldots] \quad (47)$$

The analysis for calculating channel capacity under the de-synchronization attack shows that the effect of de-synchronization is similar to increasing HWR when the channel is transparent. Increasing HWR is the factor that causes the decrease of C.

B. Simulation Results

In this section the reduction in the channel capacity by increasing δ is investigated. The analysis in the preceding section shows that the effect of the delay in de-synchronization attack is equivalent to the increase of HWR by the multiplier μ in (46). Here, the decrease of the rate of the channel capacity by the reduction in the cross correlation between W_{ji} and P and consequently the increase of HWR are examined. The channel capacity of SSW versus the increase of delay is illustrated in Fig. 6. The capacity is calculated by the parameters as: $N = 1023$, $f_x = 44100$ sym/sec (for a music signal) and different values of HWR as $HWR = 40, 60, 80, 100$. The Figure shows that the degrading rate of channel capacity would drop with the decrease of HWR. The Figure shows that the reduction of channel capacity is very small when $\delta < 0.3$.

VI. CONCLUSIONS

In this paper, the channel capacity analysis of SSW system for transmission of embedded information through a watermark channel without any attack, additive noise attack and de-synchronization attack is performed. Computing the channel capacity the highest embedding rate of information (capacity of SSW) that can attain the high reliability at the detection (robustness) is determined.

Since the paper is theoretical in nature, all the capacity formulas for different watermark channels have been proved mathematically and each Figure and Table is based on a package of the practical numbered parameters (speech or music signal) to show just a symbolic version of the proved formulas.

The channel capacity analysis of SSW shows that f_x, N and HWR are effective parameters. The effect of increasing each parameter is examined and the best period of P is derived.

We proved that Shannon's channel capacity formula for deriving the channel capacity of SSW is not directly applicable to the SSW system under additive noise attack. Instead, calculating by the channel capacity should be performed with additive color noise and using water-filling approach. Since the channel capacity analysis of color noise depends on the spectrum of noise signal, deriving a unique formula for calculating channel capacity for all audio signals is not practically possible.

Finally, calculating the channel capacity of SSW under de-synchronization attack has been performed. De-synchronization channel modeling shows that the effect of this kind of attack is similar to the increase of HWR or additive noise. In this paper, the analysis is performed by using of the increase of HWR and the decrease of the channel capacity is illustrated.

APPENDIX

The channel capacity of SSW as (32) shows that a general formula like (25) could not be derived and calculating the channel capacity completely depends on the spectrum of X. However, with the assumption that $Pow(X_i)$ varies very slowly, V/a_i remains constant and (48) will be applicable for calculating the channel capacity

$$C = \frac{f_x}{2N}\log\left[\frac{2N^2}{(HWR+\frac{1}{WNR})f_x}\right] \qquad (48)$$

Hence, using this assumption and also using (48) a general insight regarding the channel capacity of SSW under the additive noise could be obtained.

REFERENCES

[1] N. Cvejic and T. Seppänen, *Algorithms for Audio Watermarking and Steganography*, Ph.D. Dissertation, University of Oulu, 2004.

[2] M. Costa, "Writing on dirty paper," *IEEE Trans. on Information Theory*, vol. 29, no. 3, pp. 439-441, May 1983.

[3] A. Sequeira and D. Kundur, "Communication and Information Theory in Watermarking: A Survey," in *Proc. Conf. SPIE*, vol. 4518, pp. 216-227, Denver, Colorado, Aug. 2001.

[4] S. Voloshynovskiy, O. Koval, F. Pèrez-Gonzàlez, M. K. Mihcak, J. E. Vila-Forcen, and T. Pun, "Data-hiding with partially available side information," in *Proc. 13th European Signal Processing Conf., EUSIPCO*, 4 pp., Antalya, Turkry, 4-8 Sep. 2005.

[5] O. Koval, S. Voloshynovskiy, F. Deguillaume, F. Perez Gonzalez, and Thierry Pun, "Spread spectrum watermarking for real images: is everything so hopeless?" in *Proc. 12th European Signal Processing Conf.*, 4 pp., Vienna, Austria, 11-12 Sep. 2004.

[6] M. Barni, F. Bartolini, A. De Rosa, and A. Piva, "Capacity of full frame DCT image watermarks," *IEEE Trans. on Image Processing*, vol. 9, no. 8, pp. 1450-1455, Aug. 2000.

[7] M. Barni, F. Bartolini, A. De Rosa, and A. Piva, "Capacity of the watermark-channel: how many bits can be hidden within a digital image?," in *Proc. SPIE, Security and Watermarking of Multimedia Contents*, E. J. Delp and P. W. Wong, eds., vol. 3657, pp. 437-448, Jan. 1999.

[8] P. Moulin and J. A. O'Sullivan, "Information-theoretic analysis of watermarking," in *Proc. IEEE Conf., ICASSP'00*, vol. 6, pp. 3630-3633, Istanbul, Turkey, 5-9 Jun. 2000.

[9] P. Moulin and M. K. Mihcak, "The parallel-Gaussian watermarking game," *IEEE Trans. on Information Theory*, vol. 50, no. 2, pp. 272-289, Feb. 2004.

[10] P. Moulin and J. O. Sullivan, "Information theoretic analysis of information hiding," *IEEE Trans. on Information Theory*, vol. 49, no. 3, pp. 563-593, Mar. 2003..

[11] J. A. O'Sullivan, P. Moulin, and J. M. Ettinger, "Information-theoretic analysis of steganography," in *Proc. IEEE Sym. on Information Theory*, p. 297, Boston, MA, US, Aug. 1999.

[12] A. Cohen and A. Lapidoth, "The Gaussian watermarking game," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1639-1667, Jun. 2002.

[13] A. Zaidi, R. Boyer, and P. Duhmel, "Audio watermarking under desynchronization and additive noise attack," *IEEE Trans. on Information Signal Processing*, vol. 54, no. 2, pp. 570-584, Feb. 2006.

[14] C. Fei, D. Kundur, and R. H. Kwong, "The Choice of watermark domain in the presence of compression," in *Proc. IEEE Int. Conf. on Information Technology: Coding and Computing*, pp. 79-84, 2-4 Apr. 2001.

[15] I. Cox, J. Kilian, T. Leighton, and T. Shamoon, *Secure Spread Spectrum Watermarking for Multimedia*, NEC Technical Report 95-10, Princeton, NJ, US, 1995.

[16] I. J. Cox, M. L. Miller, and A. L. Mckellips, "Watermarking as Communications with Side Information," *Proceedings of the IEEE*, vol. 87, no. 7, pp. 1127-1141, Jul. 1999.

[17] Z. Liu, Y. Kobayashi, S. Sawato, and A. Inoue, "A Robust audio watermarking method using sine function patterns based on pseudo-random sequences," in *Proc. Conf. Pacific Rim Workshop on Digital Steganography*, pp. 167-173, Hsinchu, Taiwan, 2002.

[18] S. Arora and S. Emmanuel, "Real-time adaptive speech watermarking scheme for mobile applications," in *Proc. IEEE on signal processing*, vol. 2, pp. 1153-1157, Dec. 2003.

[19] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, 1991.

[20] D. Kundur, "Water-filling for watermarking?," in *Proc. IEEE Int. Conf. On Multimedia and Expo*, vol. 3, pp. 1287-1290, Aug. 2000.

[21] V. Licks, F. Ourique, R. Jordan, and F. Pérez-González, "The effect of the random jitter attack on the bit error rate, performance of spatial domain image watermarking," in *Proc. IEEE Int. Conf. Image Processing, ICIP'2003*, vol. 3, pp. 455-458, 14-17 Sep. 2003.

[22] S. Pateux, G. L. Guelvouit, and J. Delhumeau, "Capacity of data-hiding system subject to desynchronization," *SPIE*, vol. 5306, no. 28, pp. 286-295, San Jose, CA, Jan. 2004.

[23] A. Piva, M. Barni, F. Bartolini, and V. capellini, "Threshold selection for correlation based watermark detection," in *Proc. Eur. Signal Processing Conf., EUSPICO*, pp. 337-355, Sep. 1998.

M. Khademi was born in Iran, in 1958. He received the B.Sc. and M.S. degrees from Isfahan University of Technology, Isfahan, Iran, in 1985 and 1987, respectively, and the Ph.D. degree from the University of Wollongong, Wollongong, Australia, on video communications in 1995, all in Electrical Engineering. He joined Ferdowsi University of Mashhad, Iran in 1987. He is currently associated professor at the Department of Electrical Engineering, Ferdowsi University of Mashhad. Since then, he has co-chaired two conferences on "Electrical Engineering (ICEE2004)" and "Machine Vision and Image Processing (MVIP2006)" in Iran. He has received two awards including: Outstanding Graduate Student Award in 1999, and The Best Translation Award for translation of "Digital Image Processing by Gonzales" from Amir Kabir University, Iran, in 2005. His current research interests are in the areas of video communications, biomedical signal processing and information theory. He has published over 70 articles in these research fields.