Joint Data Hiding and Source Coding with Partially Available Side Information

Çağatay Dikici†, Khalid Idrissi and Atilla Baskurt

INSA de Lyon,
Laboratoire d'InfoRmatique en Images et Systèmes d'information,
LIRIS, UMR 5205 CNRS, France

ABSTRACT

Channel Coding with Side Information at the encoder(CCSI) can be visualized as a blind watermarking problem: the original host signal for embedding the watermark is known at the encoder but not at the decoder. Similarly, the Rate Distortion with Side Information at the decoder(RDSI) is known as distributed source coding: the rate distortion limits of an input source if a noisy observation of that source is available only at the decoder. There is a strong duality between CCSI and RDSI for the gaussian case.

We propose a system that exploits the generalized versions of the two information theoretical dualities of CCSI and RDSI together within a unique setup. The question is "Can we combine these two separated dual problems (blind watermarking and distributed source coding) within a single problem?". The proposed scheme can be viewed as "Watermarking or Data Hiding within Distributed Source Coding". The setup contains the cascade of the generalized versions of CCSI and RDSI where there exists two different side information, one available only at the encoder and the other at the decoder. The preliminary experimental results are given using the theoretical findings of the duality problem.

1. INTRODUCTION

The duality between the properties of a source with a distortion measure and those of a channel is stated at Shannon's landmark paper in 1959.[1] Hence the limits of channel coding and data compression are known to be dual since decades. While in the case of data compression, the rate distortion function is the minimum rate R under a distortion constraint D; for channel coding, the capacity is the maximum transmission rate within a communication channel with error probability P_e approaching zero.

The theory limits of source coding with side information at the decoder is given by Slepian and Wolf,[2] and the rate distortion limits for the memoryless gaussian lossy case is given in.[3] Moreover the channel capacity with side information at the encoder results are presented by Costa and Pinsker et al.[4,5] Source coding with side information at the decoder is known as Distributed Source Coding and the dual problem of channel coding with side information at the decoder is a digital watermarking problem. Recently, the side information duality for discrete memoryless sources and gaussian cases are demonstrated by Ramchandran et al[6,7] and Girod et al.[8] The theoretical limits of a generalized version of this duality is given by Cover and Chiang in [9] where the state information is available to the sender, receiver, to both or to neither. So all of the eight cases are generalized and the duality between them are stated.

In this paper, we look at the watermarking problem of a distributed source coding system shown in Fig. 1. Actually our solution to this problem contains the combination of the generalized version of the duality given in.[9] From the nature of Distributed Source Coding, the aim is achieving the minimum data rate for coding an input source less then a fidelity criterion D, given Y, a noisy observation of the source available at the decoder only with i.i.d. $\sim p(x, y)$. In addition to this setup, we embed a digital watermark M to the input source X with a distortion constraint between the input source and the watermarked embedded signal W such that $E[(X-W)^2] \leq D_1$. The watermarked embedded signal W is compressed to a data rate $R(W)$ such that it can be

Corresponding author currently at † INSA de Lyon, Laboratoire d'InfoRmatique, en Images et Systèmes d'information (LIRIS), Bât. St-Exupéry, 69621 Villeurbanne Cedex, France. Email: cagatay.dikici@liris.cnrs.fr.

decoded with a fixed distortion $E[(W-\hat{W})^2] \leq D_2$, given that the encoder has an access to the original data X, and decoder has an access to the noisy observation Y. The proposed system can be viewed as a quantization of the input signal X in the sense of embedding a watermark M as a function of X (or known as context dependent watermarking). The watermarked signal W is compressed with a syndrome coding or coloring for the distributed source coding. In the decoder side, the received color indexes or syndromes are decoded with the help of the side information Y, and afterwards the embedded watermark $\hat{M}$ is estimated by using $\hat{W}$, the output of the syndromes decoding and the side information Y available at the decoder.

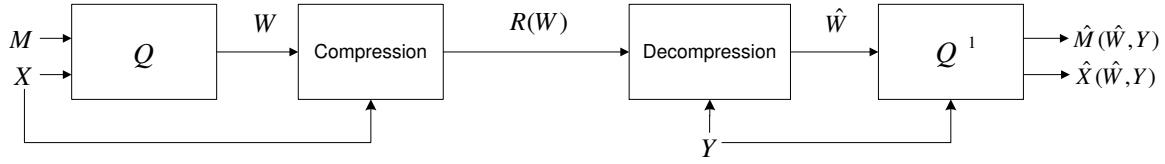

Figure 1. Data Hiding + Distributed Source Coding Scheme.

Data Hiding of Distributed Source Coding can be used on several application scenarios, from image and video coding to sensor networks. For example, in the case of encoding correlated observations of low-power sensor networks, Distributed Source Coding principles fit best for achieving power constraints. Moreover in this system, each sensor can easily hide its own hidden data within the observed signal and be coded with distributed source coding principles. This hidden data could be also served for digital rights management or contains could additional information of the sensor like the coordinates of the camera or region of interest information.

In Section.2, the duality between rate distortion and channel capacity with state information is given, both for a unique side information and the general case where there exists two different side information. A hybrid scheme is proposed for data hiding within a Distributed Source Coding setup in Section.3. Finally a preliminary simulation results of the proposed system are given in Section.4.

2. DUALITY

2.1. Rate Distortion with Side Information only available at the Decoder ($RDSI_{01}$)

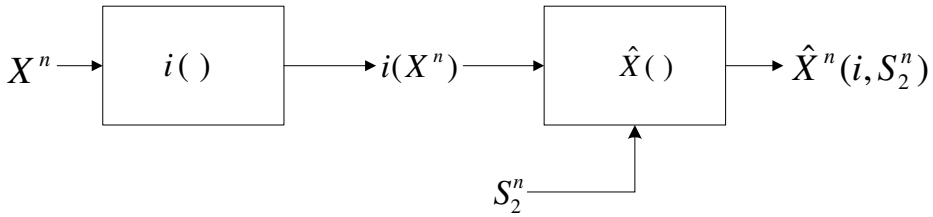

Figure 2. Rate distortion with side information available at the decoder.

Rate Distortion with side information available at the decoder is shown schematically in Fig. 2. The notation in[9] is used such that subindex 01 in $RDSI_{01}$ indicates the availability of a state information at the decoder but not the encoder. Let $\{(X_k, S_{2_k})\}$ i.i.d. $\sim p(x, s_2)$ be a sequence of independent drawings of jointly distributed random variables X and S_2. X_k is encoded with block length n into a binary stream of rate R, by using a sequence of $(2^{nR}, n)$ codes with $i : X^n \rightarrow \{1, 2, .., 2^{nR}\}$ and $\hat{X}^n : \{1, 2, .., 2^{nR}\} \rightarrow \hat{X}^n$. The input source X is to be encoded and transmitted to a receiver which access to a noisy observation S_2, and $\hat{X}$ is estimated with a fidelity criterion D such that $E[(X, \hat{X})^2] \leq D$. The minimum rate of encoding[3] for a given fidelity criteria D is:

$$R_{01}(D) = \min_{\hat{X}=f(U;S_2),p(u|x)} [I(U;X) - I(U;S_2)] \tag{1}$$

where the minimization is over all conditional probability density functions $p(u|x)$ and a function $f(U; S_2)$ such that $E(X - \hat{X})^2 \leq D$. U is defined as an auxiliary variable for the set of codewords representing X and $I(U; X)$ is the mutual information between U and X.

2.2. Channel Coding with Side Information only available at the Encoder ($CCSI_{10}$)

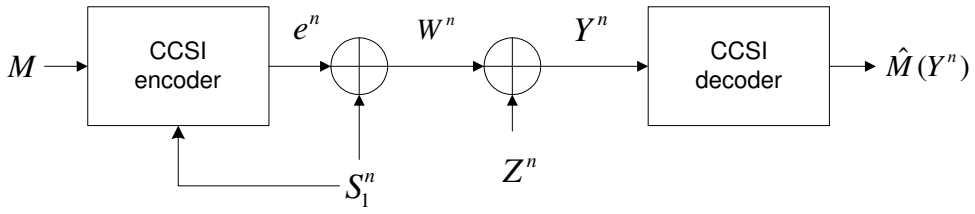

Figure 3. Channel coding with side information available at the encoder.

The blind watermarking problem can be viewed as channel coding with side information at the encoder which is shown in Fig 3. The encoder has access to a discrete watermark signal to be embedded M, and the source signal S_1 that the information is to be embedded in. There is a fixed distortion constraint between the source signal S_1 and the watermarked signal W such that $E(S_1 - W)^2 \leq D_1$. Since $W = S_1 + e$, and the error e is dependent on the input source S_1 and M, the information to be hidden, this setup is also known as content dependent data hiding. Then, the watermark embedded signal W is subjected to a fixed distortion attack Z. The achievable capacity[5] of the watermarking system for an error probability $P_e^n = Pr\{\hat{M}(Y^n, S_2^n) \neq M\}$ is:

$$C_{10} = \max_{p(u,w|s_1)} [I(U;Y) - I(U;S_1)] \tag{2}$$

where U is an auxiliary variable and the maximization is over all conditional probability density function $p(u, w|s_1)$ and $I(U;Y)$ is the mutual information between U and Y. A rate R is achievable if there exists a sequence of $(2^{nR}, n)$ codes with $P_e^n \to 0$.[9]

2.3. General Version of Rate Distortion with State Information ($RDSI_{general}$)

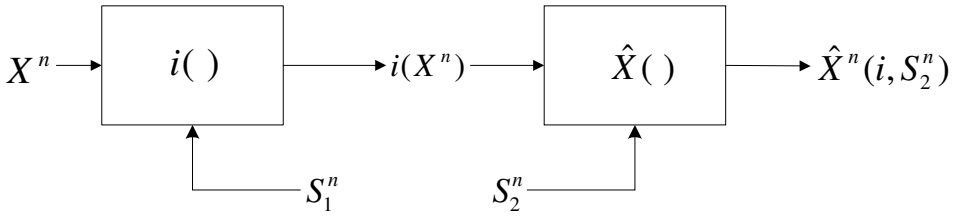

Figure 4. General version of rate distortion with two state information, one available at the encoder and the other at the decoder.

Let X, S_1, S_2 jointly distributed random variables with i.i.d. $\sim p(x, s_1, s_2)$. The setup for general version of rate distortion with state information in Fig. 4 is similar in Section 2.1, but the encoder has access to S_1, a noisy observation of X, where the decoder has access to S_2, another noisy observation of X. Then the minimum rate for achieving a fidelity criterion D is:

$$R_{S_1,S_2}(D) = \min_{p(u|x,s_1)p(\hat{x}|u,s_2)} [I(U; S_1, X) - I(U; S_2)] \tag{3}$$

where the minimization is under the distortion constraint $\sum_{x,u,s_1,s_2,\hat{x}} d(x,\hat{x})p(x,s_1,s_2)p(u|x,s_1)p(\hat{x}|u,s_2) \leq D$.[9]

2.4. General Version of Channel Coding with State Information ($CCSI_{general}$)

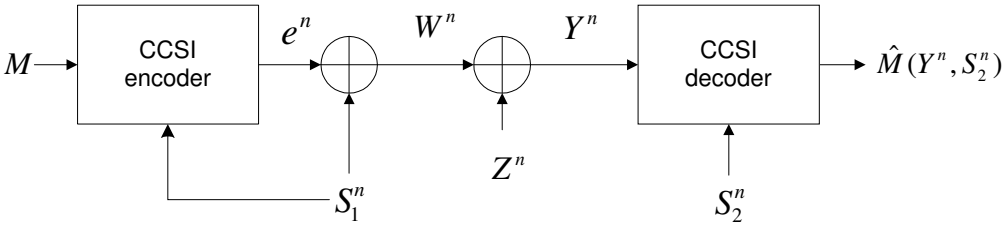

Figure 5. General version of channel coding with two state information, one available at the encoder and the other at the decoder.

The theorem given in [9] states that the memoryless channel $p(y|w, s_1, s_2)$ with state information $(S_{1,i}, S_{2,i})$ i.i.d. $\sim p(s_1, s_2)$, with S_1^n available only at the encoder, and S_2^n available only at the decoder, which can be seen in Fig.5, has capacity:

$$C_{S_1,S_2} = \max_{p(u,w|s_1)} [I(U; S_2, Y) - I(U; S_1)] \tag{4}$$

3. A GENERALIZED HYBRID SCHEME: DATA HIDING WITHIN DISTRIBUTED SOURCE CODING SYSTEM

In this section, we propose a hybrid scheme which utilize both channel coding and rate distortion with state information at the encoder and decoder respectively. The proposed system can be seen in Figure 6. Actually the system enables to hide the data M within a input source signal S_1 with a distortion measure D_1. Then the watermarked embedded signal W is compressed and transmitted with a fidelity criterion D_2 to a receiver which has access to a noisy observation S_2. Hence the encoder has access to two sources, the data to be hide or

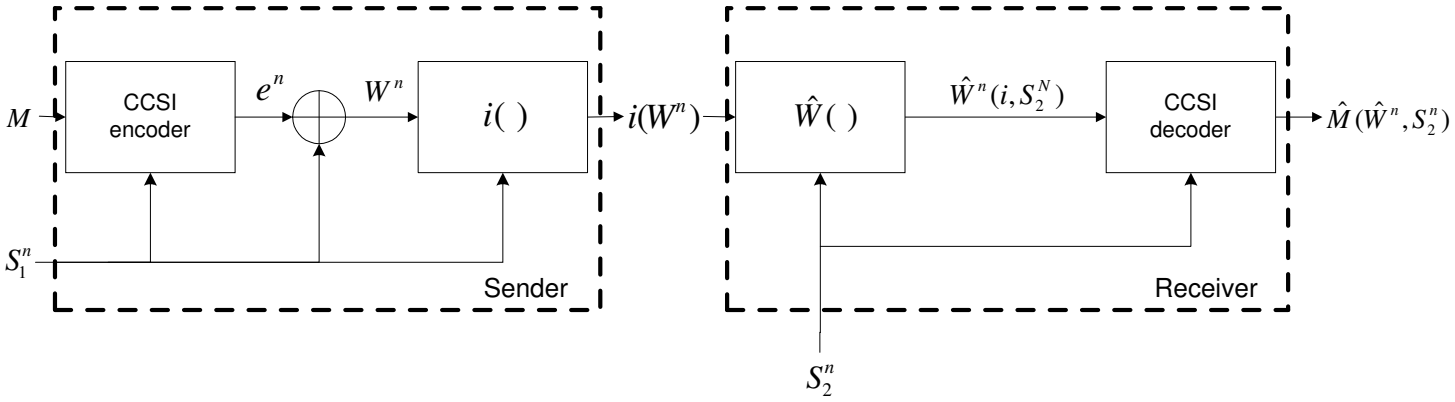

Figure 6. Proposed hybrid scheme: Data hiding within a Distributed Source Coding System.

watermark index M and the input data source S_1 that the information will be embedded in. The first criterion is embedding the watermark to the input source S_1 with a fixed-distortion measure D_1 where $E[(S_1 - W)^2] \leq D_1$. Afterwards, for a minimum rate R, the embedded watermark signal W is coded and transmitted to a receiver, given that the encoder has access to the input source S_1, while decoder has an access to S_2, which is a noisy observation of the input source S_1. The receiver decodes the received signal with the help of noisy observation S_2 with a fidelity criterion D_2 such that $E[(\hat{W} - S_1)^2] \leq D_2$ and estimates the watermarked signal $\hat{M}$ with an error probability $P_e(\hat{M})$

Mathematically, the goal is to solve the following constrained problem:

$$min_{E[(S_1-W)^2]\leq D_1, E[(\hat{W}-W)^2]\leq D_2} P_e(\hat{M}) \tag{5}$$

where $P_e(\hat{M})$ represents the decoding error probability $Pr\{(\hat{M}(\hat{W}, S_2) \neq M)\}$, and W, S_1, S_2 are jointly distributed random variables with i.i.d. $p(w, s_1, s_2)$. Moreover the distortion constraint $E[(\hat{W} - W)^2] \leq D_2$ leads to a minimum rate function:

$$R(D_2) = \min_{p(u|w,s_1)p(\hat{w}|u,s_2)} [I(U; S_1, W) - I(U; S_2)] \tag{6}$$

which is given in Equation 2.3.

In fact the proposed system is closely related to the channel coding with state information described in Section.2.4, where only the channel attack Z in Fig.5 is replaced by a fixed distortion measure created by the distributed source coding scheme with two side information, one available only at the encoder, S_1 which is the input signal that the watermark will be embedded, and one at the decoder, S_2, a noisy observation of S_1 with i.i.d. $\sim p(s_1, s_2)$.

The hybrid problem can be posed as a kind of semi-blind watermarking scheme. The receiver has not access to the input source signal S_1 in order to extract $\hat{M}$ from the watermarked signal W, but S_2, a noisy observation of the input source S_1.

4. EXPERIMENTAL SETUP AND RESULTS

Up to this point, our focus has been on the theoretical aspects of the mixture of two problems Channel Coding with Side Information and Source Coding with Side Information. In this section, we consider a real system that implements the data hiding and compression codes in joint manner. We discuss the details of the creation of codes in our system and give preliminary results.

4.1. Generation of the Side Information and Hidden Message

For our computer simulation, we used synthetically generated 1-D i.i.d. string of binary streams for side information S_1 and S_2 available only at the encoder and decoder respectively; and hidden message M to be embedded. The construction of S_1 is: i.i.d. pseudo-random Bernoulli(1/2) string of appropriate block length so the first order-entropy of $H(S_1) = 1$bit/sample. The side information available at the decoder $S_2 : S_1 \oplus N$ where the correlation noise level N between S_1 and S_2 is pseudo-random Bernoulli(p) string of the same length of the side information and $\oplus$ is the modulo-2 addition operator. The variable $p : 0 \leq p \leq 1$ controls the correlation level between two side information such that conditional entropy $H(S_1 \mid S_2) = H(p)$. And finally the hidden message M is a random binary string.

4.2. Data Hiding

For the case of informed data hiding of M within S_1, we used basic quantization based on memoryless coset construction. The algorithm is described as follows: 3 bits information is partitioned into 4 cosets such that each element of the coset has a hamming distance of 3. According to the two bits data of M the coset members of that index is chosen $Coset00 = \{000, 111\}$, $Coset01 = \{001, 110\}$, $Coset10 = \{010, 101\}$, $Coset11 = \{011, 100\}$. After creating the codebook, 2bits of M and R bits of S_1 is taken. And the least significant 3 bits of the sub-block of the host signal S_1 is depicted for embedding. The 3 bits value of S_1 is quantized to $W : W(S_1, M) = \arg\min_{X \in CosetM} \parallel X - S_1 \parallel$ which W is at most one bit differ from S_1. The distance metric is chosen as hamming distance. And this insertion of 2 bits within block length R continues until embedding all the data. As an example, assume that the 2 bits length message 01 is being embedded into the least 3 significant bits of S_1 which is 010. The minimum hamming distance between 010 and the elements of $Coset01$ is chosen as the quantification output, which is $W = 110$ in this case. At the decoder side, the extraction of the watermark is straightforward such that the knowing the codebook and insertion frequency R, the coset index that the received block data resides in is decoded as the embedded data.

4.3. Compression and Decompression Setup

The watermarked string W is compressed by finding its syndrome with respect to a low-density parity check(LDPC) channel code.[15] In fact we use high-rate (3/5) LDPC code and transmit only the check bits to the encoder. A code expressed as (n,k) where $m = n - k$ check bits are calculated from input stream of length k using a randomly generated parity check matrix H with dimension $(n-k) \times n$ whose codewords satisfy $Hx = 0$. Since only the parity check bits are sent to the decoder, so the compression rate is the number of check bits over input length $(n-k)/k$.

At the receiver, the compressed data is decoded using belief propagation algorithm in,[14,15] with the help of side information S_2 available only at the decoder. The goal of decoding is to find the nearest likelihood codeword $\hat{W}$ and extract the embedded string estimation $\hat{M}$. The side information is assumed to be the systematic bits and the received compressed data is assumed to be the parity checks. The belief propagation algorithm is very identical to that used for decoding standard LDPC codes, with some modifications to in our case. First, likelihood ratios of the systematic bits are initialized according to the correlation noise N between the two side information S_1 and S_2. Second, initial likelihood ratios of the parity check bits are based on the fact that probability of received parity check is in error with a small probability ε. Moreover check-node update node is modified to recover the errors on the systematic bits using the fact that check-bits are correct with high probability. Finally, with the knowledge of the coset codebook and estimation of $\hat{W}$ using LDPC decoding its trivial to extract the hidden data $\hat{M}$. The distortion levels of the $\hat{W}$ and $\hat{M}$ are given in results.

4.4. Simulation Results

In our experiments, 100 blocks each of 2000bit length of host signal S_1 is generated. With a correlation noise ratio of $p : 0 < p < 0.2$, the side information S_2 available only at the decoder is created as in Sec.4.1. In the first set of experiments, we tested the performance of the compression system without embedding a watermark. The S_1 is compressed with 3/5 rate LDPC (3500,2000) code as explained in Sec.4.3. Hence for each block, the 1500 parity check bits are sent to the receiver. And receiver decodes the received parity check bits using S_2 the side information available at the decoder, using belief propagation with maximum of 50 iterations. Afterwards, in the same setup, a 4000 bits length hidden message is embedded while changing the correlation noise between side information. In the second set of our experiments, for each R bits data of S_1, 2 bits long of a hidden message is embedded as described in Sec.4.2. R is varied between 50 and 2000. With a fixed correlation noise $p = 0.1$ between the side information, the signal to embedded watermark noise ratios between host data S_1 and watermarked data W are given for various length of embedded signal.

As seen in Fig.7, for the case of our distributed source coding scheme, up to a correlation noise error of $p < 0.13$ the compressed host signal can be obtained without error. Please note that the theoretical limits of the probability of error for channel capacity of $C = 3/5$ is $p_e = 0.08$ over all transmitted bits. Since the correlation noise ratio p is defined only on the systematic bits, the overall achieved ratio of our system is $p_e = 0.13 * 3/5 = 0.78$, close to the theoretical limits. The results of the same experimental setup with embedding 4000 bits length of hidden data M is also given in Fig 7, where the the error between the side information S_1 and the estimated watermarked signal $\hat{W}$ at the decoder is plotted. Recall that for $p < 0.13$, all of the 4000 hidden bits are are extracted wperfectly. And we have a hidden message error rate of $p_{hm} = 1.75 \times 10^{-3}$ for $p = 0.13$, $p_{hm} = 7.5 \times 10^{-3}$ for $p = 0.14$, $p_{hm} = 5.2 \times 10^{-2}$ for $p = 0.15$ and $p_{hm} = 1.24 \times 10^{-1}$ for $p = 0.16$. For the results of the second set of experiments, a fix correlation noise error of $p = 0.1$ is used for errorless watermark extraction. Below are some of the signal to watermark power as a function of Embedded bit size: $E_b/N_w = 34.32dB$ for 200 bits of watermark, $31.08dB$ for 400 bits, $28.30dB$ for 800 bits, $21.16dB$ for 4000 bits.

5. CONCLUSIONS

In conclusion, we establish a hybrid system for hiding data to a compression process which uses distributed source coding system. Recent findings about the duality between rate distortion and channel capacity with state information are used for the system. The hybrid scheme proposed in Fig.6 offers a wide range of multi-source coding systems. The selection of the inputs such as the side information S_i of both sides and the nature of the hidden information M depends on the considered problem. A memoryless data hiding algorithm is used with LDPC based distributed compression scheme. A trellis based data hiding with memory can be used for improving

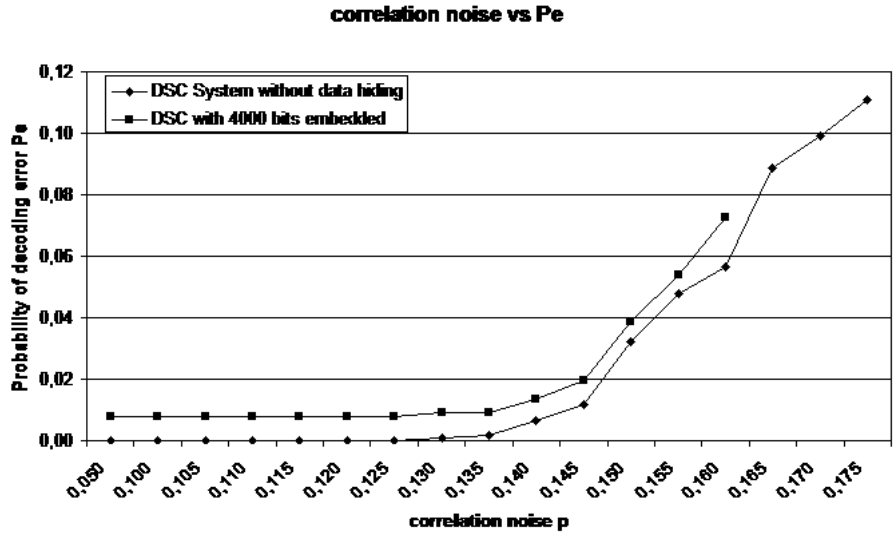

Figure 7. Simulation results.

the performance of the overall system. Indeed, this scheme can be easily adapted for example to video coding such that temporally correlated successive frames serve as sources S_i to be coded and the watermark signal to be embedded serves for M.

REFERENCES

1. C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion", In IRE Nat. Conv. Rec., vol. Part 4, pp. 142–163, 1959.
2. J. D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources", *IEEE Transactions on Information Theory*, vol. IT-19, pp. 471–480, July 1973.
3. A. D. Wyner and J. Ziv, "The Rate-Distortion Function for Source Coding with Side Information at the Decoder", *IEEE Transactions on Information Theory*, vol. IT-22, no. 1, pp. 110, Jan. 1976.
4. M. Costa, "Writing on dirty paper," IEEE Trans. on Information Theory, vol. 29, pp. 439–441, May 1983.
5. S. Gel'fand and M. Pinsker, "Coding for channel with random parameters," Problems of Control and Information Theory, vol. 9, pp. 19–31, 1980.
6. J. Chou, S. S. Pradhan, and K. Ramchandran, "On the duality between distributed source coding and data hiding," Proc. of 33rd Asilomar Conf. on Signals, Systems and Computers, November 1999.
7. S. S. Pradhan, J. Chou and K. Ramchandran. "Duality between source coding and channel coding and its extension to the side information case." IEEE Transactions on Information Theory, vol. 49, vo. 5, May 2003. IEEE, USA.
8. J. K. Su, J. J. Eggers and B. Girod, "Illustration of the Duality Between Channel Coding and Rate Distortion with Side Information," 34th Asilomar Conf. on Signals, Systems, and Computers. Oct. 29-Nov. 1, 2000, Asilomar, CA, USA.
9. T. M. Cover and M. Chiang, 'Duality between channel capacity and rate distortion with two-sided state information,' IEEE Trans. of Inform. Theory, vol. 48, no. 6, pp. 1629 - 1638, June 2002.
10. B. Rimoldi and R. Urbanke, "Asynchronous SlepianWolf coding via sourcesplitting," Proc. IEEE Int. Symp. on Info. Theory, Ulm, Germany, July 1997.
11. S. S. Pradhan and K. Ramchandran, "Distributed Source Coding: Symetric rates and applications to sensor networks.",*Proc. IEEE Data Compression Conf.*, pp.363-72. Los Alamitos, 2000.
12. A. Aaron and B. Girod, "Compression with side information using turbo codes", *Proc. IEEE Data Compression Conf.*, pp. 252–261, 2002.
13. T.J. Flynn and R.M. Gray, "Encoding of correlated observations", *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 773–787, Nov. 1987.
14. R. G. Gallager, Low density parity check codes, Ph.D. dissertation, MIT, Cambridge, MA, 1963.
15. MacKay, D. J. C. and R.M. Neal,"Near Shannon limit performance of low density parity check codes",*Electronics Letters,* vol. 33, pp. 457-458, 1996.