Efficient Codebook Structures for Practical Information Hiding Systems

H. T. Sencar
Electrical and Computer Engineering Department
New Jersey Institute of Technology
University Heights, Newark, NJ 07102-1982
e-mail: ths1034@njit.edu

M. Ramkumar
AVWAY.COM Inc.
IDT Corporation
520 Broad Street, Newark, NJ 07102
e-mail: ramkumar@corp.idt.net

A. N. Akansu
Electrical and Computer Engineering Department
New Jersey Institute of Technology & NJCMR
University Heights, Newark, NJ 07102-1982
e-mail: ali@megahertz.njit.edu

Abstract — **Conventional information hiding systems suffer from communicating at low rates. Generating efficient codebooks and codewords for information embedding are of crucial importance for reliable communication. In this paper, we present new insights and interpretations to the problem of efficient information hiding, by improving codebooks utilized by practical information hiding schemes. We also explore codeword selection and optimize codewords that achieve higher rates. We evaluate information hiding rates in terms of distortions introduced by information hider and attacker within the context of data hiding method employed.**

I. Introduction

As multimedia technologies go through great advances and become extensively available, issues like copyright protection and authentication become the focus of attention for multimedia content providers. Information hiding promises to address most of these concerns.

Information hiding is the art of hiding a message into a host signal or content like image, audio or video with no perceptual distortion to the host signal. In data hiding applications, the limitation for the information hider and the attacker is the maximum distortion that may be introduced to the content without affecting the fidelity of the original content.

Information hiding can be classified into two main categories based on the need for the host signal for extraction of the hidden message, namely, non oblivious and oblivious information hiding. Non oblivious hiding systems need the host signal for extraction of the message. However, in oblivious information hiding systems extractor generally has no access to the host signal. We focus our attention on oblivious information hiding systems.

Fundamentals of information hiding research was established by the work of Gel'fand and Pinsker [1], although they did not directly, address it. They derive the capacity of a discrete memoryless channel in the presence of side information known only to the decoder. Later research, gained a considerable momentum by reinterpreting side information as the multimedia content that will carry hidden information. Heegard and El Gamal [2] use this formulation and extend it to establish achievable storage rates for memory when defect information is given completely to decoder and partially to encoder.

Costa [3], was the first researcher, to present an information-theoretic analysis of a channel with side information. He evaluated channel capacity and proposed encoder-decoder structures that will achieve this capacity. Emerging efforts in this field brought Costa's paper to researcher's attention and made it a source of inspiration for designing data hiding schemes and evaluating their performances.

Moulin and O'Sullivan in [4] and Cohen and Lapidoth in [5] give complementary insights into the information-theoretic analysis of the problem and highlight the design criteria for practical systems. Chen and Ramkumar proposed methods using special quantizers in [6] and [7, 8] respectively.

In this paper, we present new insights and interpretations to the problem of efficient information hiding. It is shown that improving codebooks, utilized by practical information hiding schemes, increases data hiding rates. We also explore codeword selection and optimize codewords that achieve higher rates. We adopt a communication framework where the channel input is the watermarked content and the channel output is a distorted version of it. Attacks to the content are assumed to be additive white Gaussian noise. Our analyses evolve from the information hiding technique represented in [7]. For a better understanding of the analysis reader should refer to it. We evaluate information hiding rates in terms of the metric, defined as the ratio of the watermark power to the attack power. In the next section we present a description of the information hiding model. In sections III and IV analyses and results for an efficient information hiding system are presented.

II. Information Hiding

Consider a discrete memoryless channel with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$, both of which depend on a given side information from a finite set $\mathcal{S}$ where $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{S}$, $\in \mathcal{R}^n$. Channel capacity is expressed with the addition of an auxiliary random variable $U \in \mathcal{U}$, $\mathcal{U}$ being a finite alphabet in $\mathcal{R}^n$, for the conditional joint probability density $p(U,X|S)$, $X \in \mathcal{X}$, and $S \in \mathcal{S}$ is, as in [1],

$$R = max_{p(U,X|S)}\{I(U,Y) - I(U,S)\}. \quad (1)$$

To design $U = X + \alpha S$, under the input power constraint $\frac{1}{n}\sum_{i=1}^{i=n} X_i^2 \le P$, Costa derived the communication rate for $Z \sim \mathcal{N}(0,\sigma_n^2 I)$, $S \sim \mathcal{N}(0,\sigma_s^2 I)$, $X \sim \mathcal{N}(0,PI)$ ($\mathcal{N}(0,1)$ is a zero mean Gaussian random varible with variance one and I is identity matrix), where X, S, Z are independent, as;

$$R(\alpha) = \frac{1}{2}\ln\left(\frac{P(P+\sigma_s^2+\sigma_n^2)}{P\sigma_s^2(1-\alpha)^2+\sigma_n^2(P+\alpha^2\sigma_s^2)}\right). \quad (2)$$

Costa proved that by setting $\alpha = \frac{P}{P+\sigma_n^2}$ in (2), the channel achieves its capacity, $C = \frac{1}{2}\ln(1+P/\sigma_n^2)$. This is the capacity of an additive white Gaussian noise channel where the side information is known to both encoder and decoder. Costa also outlined the encoder-decoder structures by choosing a jointly typical (U, X, S) and (U, Y) from the codebook consisting of $e^{nI(U,Y)}$ codewords uniformly partitioned into e^{nR} bins.

A message indexed by m, from an alphabet $\mathcal{M}$, $1 \leq m \leq M$ is mapped to a sequence $W \in \mathcal{R}^n$. Sequence W must be embedded into the original content without any perceptual distortion. It is obvious that choosing X completely independent of the content, S, will not necessarily minimize distortion. This suggests more sophisticated embedding structures that minimize distortion as they are immune to attacks. However it is not easy to impose generic distortion measures for all applications. Imposing restrictions on the distortions introduced by the information hider and attacker, such that $\frac{1}{n}\sum_{i=1}^{i=n} S_i^2 \gg P$, and $\frac{1}{n}\sum_{i=1}^{i=n} S_i^2 \gg \frac{1}{n}\sum_{i=1}^{i=n} Z_i^2$, will keep the original content more or less intact and simplify the problem. Detailed discussion of small distortion analysis may be found in [4]. As the input signal is power limited, this communication model assumes squared error distance as the measure of distortions introduced by information hider. It is however worth noting that ideal distortion measure should be in compliance with the perceptual properties of the content.

For a conventional oblivious information hiding scheme, a direct projection on the Costa's framework is not possible due to two major issues: generating the optimum codebook (U) and conforming the encoder-decoder structures. The channel achieves its capacity when $\alpha = \frac{P}{P+\sigma_n^2}$ which is a function of attacker behavior, σ_n^2. This implies that ideal scheme will be designed only if information hider is given a particular prior channel knowledge. The latter is due to the size of codebooks that both encoder and decoder should retain. Ref. [9] argues that the ideal encoder-decoder requires codebooks $\approx 2^{40}$ times the size that a conventional scheme needs.

Overall information hiding system is outlined below where $m \in \mathcal{M}$ is the message to be hidden, S is the original content data, X is the distortion introduced by the hider and Z is the intrusion of the attacker:

$$\begin{aligned} \mathcal{W} &: \quad m \longrightarrow W, \\ S + X &= \mathcal{E}(S, W), \\ Y &= S + X + Z = \mathcal{E}(S, W) + Z, \\ \hat{W} &= \mathcal{D}(Y), \\ \mathcal{W}^{-1} &: \quad \hat{W} \longrightarrow \hat{m}. \end{aligned} \tag{3}$$

The mapping $\mathcal{W}$ from m to W is a one to one mapping in $\mathcal{R}^n$. The embedder, $\mathcal{E}$, and detector, $\mathcal{D}$, are nonlinear and not necessarily invertible functions $(\mathcal{D}(\mathcal{E}(S, W)) \neq W)$. Later, we show that non invertible embedder-detector set is used to increase the information hiding rate. The auxiliary random variable U, also functions as the codebook within information-hiding context. It is represented by the embedder-detector which is a dual of coder-decoder based on joint typicality in the formulation. The mapping, $\mathcal{W}$, and inverse mapping, $\mathcal{W}^{-1}$ may also be included within the embedder-detector structure. The average probability of error in detecting m from $\hat{W}$ is

$$P_e = \frac{1}{M}\sum_{m \in \mathcal{M}} P(\mathcal{D}(Y) \neq W | M = m). \tag{4}$$

In order to attain higher information-hiding rates information hider has two degrees of freedom. These are choosing the codebook and codeword. Based on the content, S, encoder will choose the best codeword, $X + \alpha S$, with the minimum distortion to the content data.

In order to maximize the rate, information hider should maximize the distortion introduced in the content. Similarly, attacker will distort the watermarked content to remove tracks of the embedded message. Communication rate is a tradeoff between the distortions of the information hider and attacker. The codebook $U = X + \alpha S$ should be designed for the information hiding scheme and under best attack assumption. In [4], it is shown that for oblivious information hiding, Gaussian attack is the optimal one. Communication rate achieves capacity when the ideal codebook $U = X + \alpha S$ is generated for $\alpha = \frac{P}{P+\sigma_n^2}$. The codebook design $U = X$ for $\alpha = 0$ suffer from dramatically low hiding rates. Because $U = X$ is a nonoptimal codebook design which assumes S as an extra noise and tries to cancel it. In applications embedder-detector set is a function of α.

Information hider must consider that watermarked content will confront attacks with high distorting power, $P < \sigma_n^2$. Under this assumption, α for the optimal codebook design may deviate far away from 0.5. Figure 1 presents the channel communication rates for 5 different codebooks, $U = X + \alpha S$, and $0 \leq P \leq 3\sigma_n^2$.

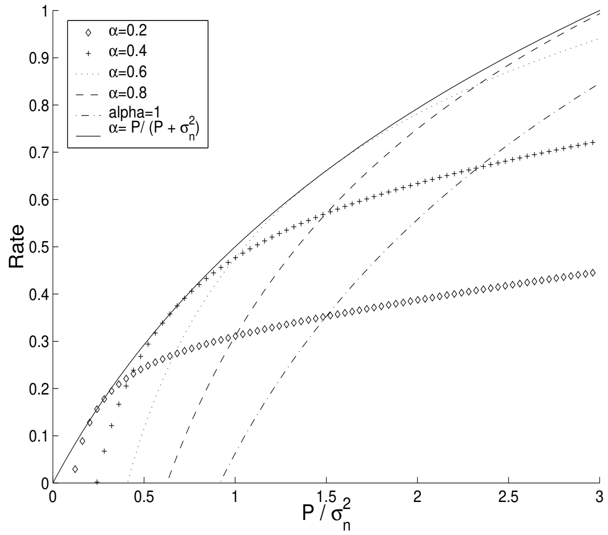

Fig. 1: Channel Rates for 5 Codebooks obtained for varying α

The codebook structure $U = X + \alpha S$ is realizable when $\alpha = 1$, by the use of special quantizers whose input is uncorrelated with the quantization error. These quantizers take S and W as the input and output $X + S$ where X and S are uncorrelated. References [6] and [8] employ embedder-detector sets using this type of quantizers. Disadvantage of choosing $U = X + S$, $\alpha = 1$, in an oblivious information hiding system is that the system performs better if $\frac{P}{\sigma_n^2} \gg 1$ because communication rate achieves the capacity only when there is no attack. Additionally, channel communication is not possible when $P \leq \sigma_n^2$ assuming $\sigma_s^2 \gg P$ and $\sigma_s^2 \gg \sigma_n^2$. However, embedder-detector sets utilizing these quantizers are limited to the codebooks generated for $\alpha = 1$ because scaling S will cause noticeable changes on the host data, most of the time.

III. Results

Information hider has the freedom to choose X via the embedding method being employed. In order to achieve higher information hiding rates distortion, X, should be optimized. Reordering the terms in (1) for $U = X+\alpha S$ and $Y = S+X+Z$, we get

$$R(\alpha) = max_{p(U,X|S)}(H(X) + H(X+S+Z) \\ -H(X+S+Z, X+\alpha S)). \quad (5)$$

Among the three terms contributing to (5), third one is less dependent on pdf of X, under the assumption that S and N are Gaussian. First and second terms are maximized when X is Gaussian distributed. So, an optimal codeword is generated when $X \sim \mathcal{N}(\mu, P)$ where μ is an arbitrary number, $X \in \mathcal{X}$ and $\mathcal{X} \in \mathcal{R}^n$. In other words, $R(\alpha)$ is maximized in codeword statistics, if the entropy $H(X)$ is maximum under the input power constraint. Based on this, the quality of an embedder-detector set mapping, a message m to the codeword X can be decided upon "distance" between X's distribution and a Gaussian distribution.

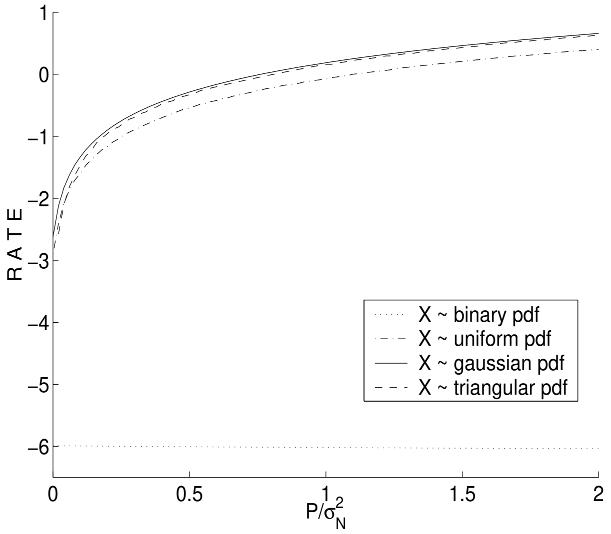

Fig. 2: R(α) obtained for different distortion statistics.

Figure 2 demonstrates $R(\alpha = 1)$, (5), calculated for various X distributions when $P \leq 2\sigma_n^2$. These results suggest that binary and uniform distributed distortion pdfs are not appropriate for practical applications. Moreover, embedders with binary distributed distortions can not be used in a reliable information hiding scheme. For $P \leq \sigma_n^2$ none of the practical schemes seem to be promising because no positive rate is possible.

In [8], Ramkumar proposed a practical scheme that improves the hiding rate by removing invertibility condition for $\mathcal{E}, \mathcal{D}$ sets. In the following analysis, we formalize the approach in [8] and show that **channel communication is possible for** $P \leq \sigma_n^2$ with the codebook $U = X + S$, i.e. $\alpha = 1$ with a sacrifice of hiding rate for $\frac{P}{\sigma_n^2} > 1$.

Ref. [8] includes modifications to the encoder structure by imposing limitations on X. As a result, distortion X corresponding to message m is replaced by X_n. Embedder derives X_n from X and imposes the input power constraint on X_n rather than on X. Because decoder is not aware of this change $X_d = X_n - X$ can be considered as another source of noise. This may also be interpreted as dependency between the total channel noise, $Z' = Z + X_d$, and hider's distortion, X, to the content.

Modified embedder-detector system may be represented as communication through a discrete memoryless channel where channel noise is correlated with the side information. Hiding rate for this scheme may be still be calculated using (1). Considering the same channel described in [3] with the modification that the channel noise, Z, and the introduced distortion, X, are correlated, $E[XZ] = L$. The rate, $\acute{R}(\alpha)$, where $L = \rho_{XN}\sigma_x\sigma_n$ is

$$\acute{R}(\alpha) = \frac{1}{2}\ln\left(\frac{A}{B}\right) \quad (6)$$

$$\begin{aligned} A &= P(P + \sigma_s^2 + \sigma_n^2 + 2L) \\ B &= P\sigma_s^2(1-\alpha)^2 + \sigma_n^2(P + \alpha^2\sigma_s^2) + \sigma_s^2 L(2\alpha^2 - 2\alpha) - L^2 \end{aligned} \quad (7)$$

For $L = 0$, $\acute{R}(\alpha)$, (7), is equal to the rate of the channel described by Costa as given in (2). $\acute{R}(\alpha)$ is maximized for $\acute{\alpha} = \frac{(P+L)}{(P+\sigma_n^2+2L)}$,

$$\acute{R}(\acute{\alpha}) = \frac{1}{2}\ln\left(\frac{P(P+\sigma_n^2+2L)}{P\sigma_n^2 - L^2}\right). \quad (8)$$

Equation (8) is the capacity of a Gaussian channel with correlated noise and input. Figures 3 and 4 represent the rate obtained for $P \leq \sigma_n^2$ at different correlation values for X and Z.

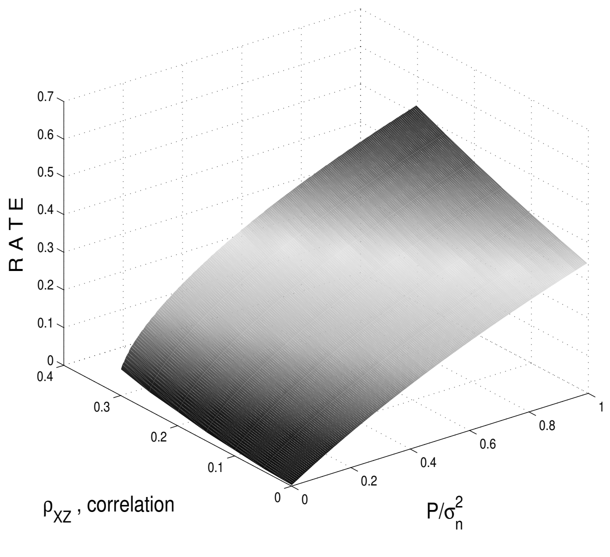

Fig. 3: Rate obtained for various ρ_{XN}.

This analysis may be extended for practical schemes by using a similar analysis as in (7) for $\alpha = 1$. Let $X_d \in \mathcal{X}$ be a deterministic function of X, with correlation $E[X, X_d] = \rho_{XX_d}$, known to the embedder but not to the detector. Therefore, X_d is also independent of Z and S. After choosing the codeword, as in the original scheme, corresponding to message m and content S from the codebook U we modify X to $X_n = X + X_d$ under the power constraint $\frac{1}{n}\sum_{i=1}^{i=n} X_{n_i}^2 \leq P$ where detector is blind to this modification.

Recalling (1) for auxiliary random variable $U = X + S$ and reordering the terms rate, $\acute{R}$, for this scheme is derived as

$$\acute{R} = I(U,Y) - I(U,S),$$

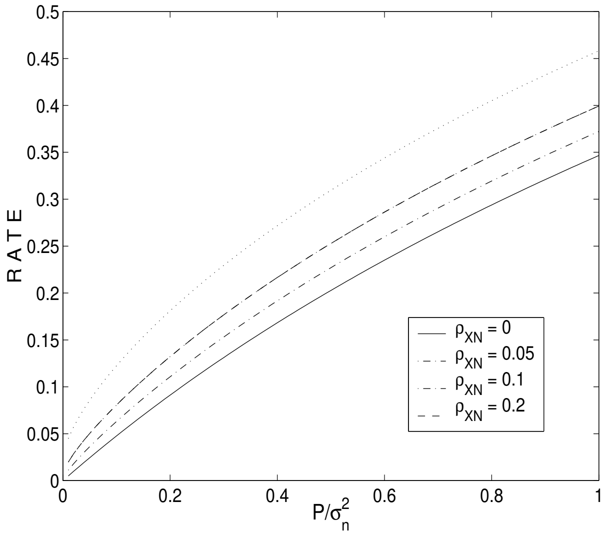

Fig. 4: Rate obtained for various ρ_{XN}.

$$\begin{aligned}
\acute{R}(\alpha=1) &= I(X+S, X+X_d+S+Z) - I(X+S,S),\\
\acute{R}(\alpha=1) &= H(X+X_d+S+Z) - H(X+S)\\
&- H(X+X_d+S+Z|X+S) + H(X+S|S),\\
\acute{R}(\alpha=1) &= H(X_n+S+Z) - H(X+S)\\
&- H(X_d+Z|X+S) + H(X),\\
\acute{R}(\alpha=1) &= H(X_n+S+Z) + H(X)\\
&- H(X_d+Z, X+S). \qquad (9)
\end{aligned}$$

Comparing this result with the rate reevaluated in (1) for $\alpha = 1$, it is shown in Figure 5 that channel rate is improved for the region $P \leq \sigma_n^2$.

$$R(\alpha=1) = H(X+S+Z) + H(X) - H(Z) - H(X+S) \quad (10)$$

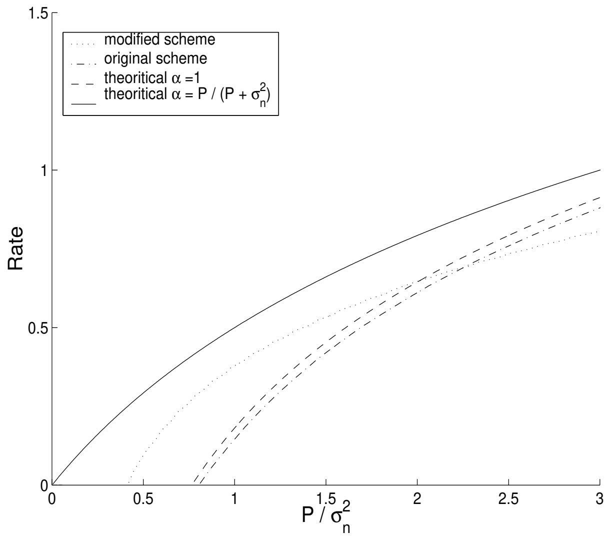

Fig. 5: Channel Rate Comparison for the Scheme

Figure 5 shows the rates for the scheme, (10), and modified scheme, (9), computed and compared to the channel rates derived for both $\alpha = 1$ and $\alpha = \frac{P}{P+\sigma_n^2}$ where P varied in $0 \leq P \leq 3\sigma_n^2$ under independent and Gaussian distributed Z and S assumptions. Hiding rate for the modified scheme is closer to the channel capacity than the rate $R(\alpha = 1)$, (2), for $P \leq \sigma_n^2$. While calculating the channel rates for the modified and original schemes the codewords are not assumed to be Gaussian distributed because there is no easy way for a practical system to generate Gaussian distributed codewords. However with a proper mapping $\mathcal{W}: \ m \longrightarrow W$ and embedder $\mathcal{E}$ it is possible to obtain distortions with triangular distribution which has entropy very close to entropy of $X \sim \mathcal{N}(0, P)$. The channel rate for the original scheme in Figure 5 slightly under performs the rate (10) for $\alpha = 1$ due to this nonoptimal codeword selection.

IV. Experiments and Conclusion

Simulation results reveal a means for comparing the original and modified schemes. We assumed an alphabet of size sixteen, $M = 16$. Each message is mapped to a row or sign inverted row of $\frac{M}{2} \times \frac{M}{2}$ Hadamard matrix, so that an orthogonal codebook of $U = X + S$, $\alpha = 1$, is generated. A vector of Gaussian distributed data representing a multimedia content is used for embedding the message. Watermarked signal undergoes an additive Gaussian noise attack. Detector extracts the message from the composite signal using a minimum distance decoder. Figure 6 displays the Probability of successfully detecting a codeword, P_s, obtained at different signature powers. The straight line, $P_s = \frac{1}{M}$, is the probability where using a detector does not help a reliable communication. Efficient decoding under this noise level is equivalent to picking any of the messages randomly as if it is sent.

Figure 7 attributes the probability of true detection to symbol rate by approximating the information hiding system with an Mary symmetric channel. The importance of a proper distortion, X, on communication rate is also demonstrated for uniform and triangular distributions. Embedders generating distortions with uniform distortion has poor communication rates.

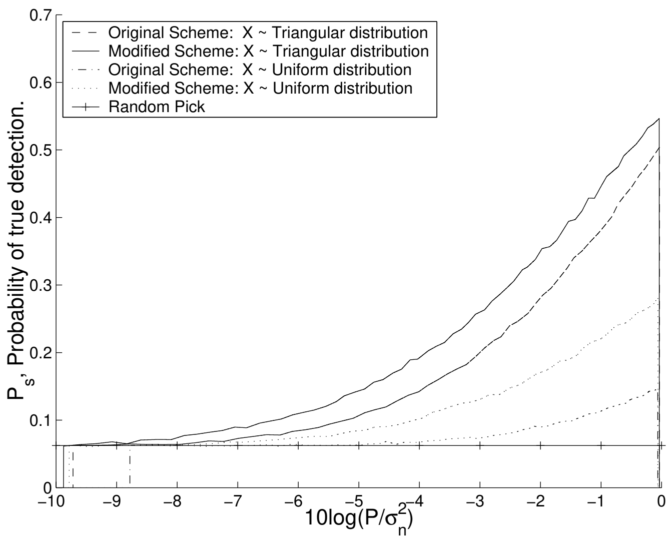

Fig. 6: Detection statistics for the described schemes.

Modified scheme can also be interpreted as the embedder whose performance may be maximized by adapting the codebook to the channel noise. Given the prior channel information (σ_n^2) ideal modified scheme performance can be evaluated. Figure 8 displays the rate for ideal modified scheme. On the two extremes where P/σ_n^2 is small and high the performance of the scheme is equivalent to using codebooks $U = X$ and

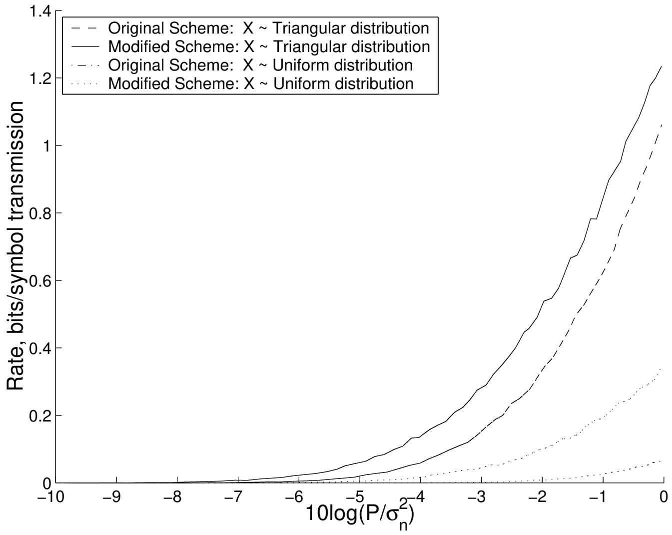

Fig. 7: Capacity under binary symmetric channel assumption.

$U = X + S$, respectively. Ideal modified scheme has the same performance of the original scheme for high P/σ_n^2 ratios. For small ratios communication, information hiding, is still possible with low rates.

These results suggest that the embedder structure given in outline (4) may be replaced by $S + X = \mathcal{E}(S, X, X_d)$ with the known correlation, $E[X, X_d] = \rho_{XX_d}$ and no change in the detector. When channel is subject to a poor attack, $P \geq \sigma_n^2$, the channel rate for modified scheme is below the original scheme. This requires original and modified schemes to be employed in parallel in a practical information hiding system.

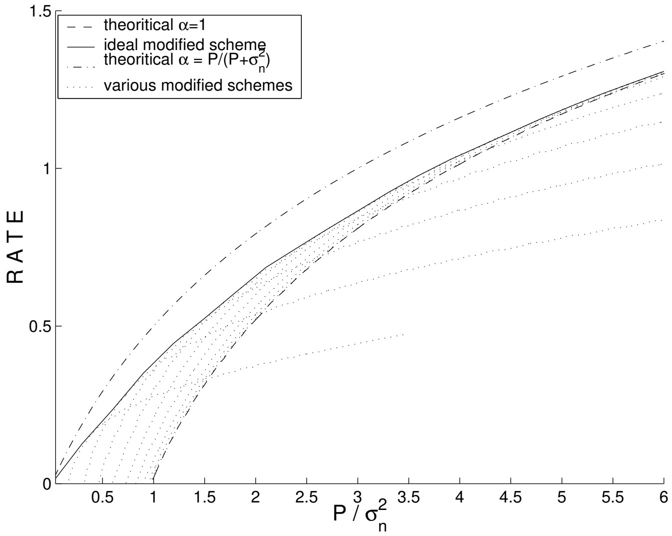

Fig. 8: Rate of the ideal modified scheme.

References

[1] S.I. Gel'fand and M.S. Pinsker, "Coding for Channel with Random Parameters," Problems of Control and Information Theory, Vol. 9, No. 1, pp 19–31, 1980.

[2] C. Heegard and A. A. El Gamal, "On the Capacity of Computer Memory with Defects," *IEEE Trans. Info. Thy*, Vol. 29, No5, pp 731–739, Sep. 1983.

[3] M. Costa, "Writing on Dirty Paper," *IEEE Transactions on IEEE Trans. Info. Thy*, Vol. 29, No. 3, pp 439–441, May 1983.

[4] P. Moulin and J.A. O'Sullivan, "Information-Theoretic Analysis of Information Hiding," *preprint*, Sep. 1999, http://www.ifp.uiuc.edu/∼ moulin/paper.html

[5] A.S. Cohen and A. Lapidoth, "The Gaussian Watermarking Game Part I-II," *preprint*, Jan. 2001, http://www.mit.edu/people/acohen/Pubs/

[6] B. Chen and G.W. Wornell, "Provably Robust Digital Watermarking," *Proc. of SPIE: Multimedia Systems and Applications II (Photonics East'99)*, Boston, MA, Vol. 3845, pp 43–54, Sep. 1999.

[7] M. Ramkumar and A. N. Akansu, "Self-Noise Suppression Schemes for Blind Image Steganography," *Proc. of SPIE: Multimedia Systems and Applications II (Photonics East'99)*, Boston, MA, Vol. 3845, pp 55–68, Sep. 1999.

[8] M. Ramkumar, A. N. Akansu, and X. Cai, "Floating Signal Constellations for Multimedia Steganography," *Proc. of IEEE ICC*, New Orleans, LA, Jun. 2000.

[9] M. Ramkumar and A. N. Akansu, "Robust Protocols for Proving Ownership of Images," *Proc. of IEEE-ITCC*, Las Vegas, NV, pp 22–27, Mar. 2000.