Research on Data Hiding Capacity

Zhensong Liao[1], Yan Huang[2], and Chisong Li[1]
(Corresponding author: Zhensong Liao)

School of Computer, HuaZhong University of Science and Technology[1]
Wuhan, West 13, Bedroom 323, P. R. China (Email: {zsliao, lizan}@mail.hust.edu.cn)
School of Control, HuaZhong University of Science and Technology[2]
Wuhan 430074, P. R. China, Wuhan (Email: huangyan2323@163.com)

(Received Nov. 09, 2005; revised and accepted Apr. 7, 2006)

Abstract

To hide data in credentials is a key problem in information security. In this paper, a summary of the work on data hiding-capacity is made and several communication channel models and several statistical host data models have been considered by the application of information theory. Based on the foundation, a solution is given to the problem about how many bits can be hidden in host data transparently and robustly.

Keywords: Authentication, credential, data hiding

1 Introduction

Data hiding refers to the general process by which a discrete information stream is embedded within a multimedia signal by imposing nearly invisible changes on the host signal, such as text, audio, image, or video. There is a variety of data one may want to hide in such data sets. The hidden information may be a textual description of image features, some complementary information (words, sound, etc.) about the original scene, or something has nothing to do with the host data. Information hiding has many application areas, such as the copyright protection for digital media watermarking, fingerprinting, steganography and data embedding. In data hiding applications, the hidden data can represent authorship information, a time stamp, or copyright information.

Mostly, information hiding method is applied when people try to transmit some information secretly, but there must be some malicious opponents apply various operations to interfere with the process, they want to get this information, or corrupt this information. Transmitter must be able to recognize the presence of an attacker who attempts to disrupt the communication of hidden data despite that information is corrupted only by un-malicious manipulation.

Our focus is on transferring significant amounts of information to a decoder. Channel capacity is the maximum data transmission rate across a communication channel with the probability of decoding error-approaching zero, and the rate distortion function is the minimum rate needed to describe a source under a distortion constraint.

2 Channels Model

Moulin and O'Sullivan have set up a channel model in [6], which has be accepted and applied so widely that it becomes a standard model. Here, we setup a channel model based on that model, shown in Figure 1.

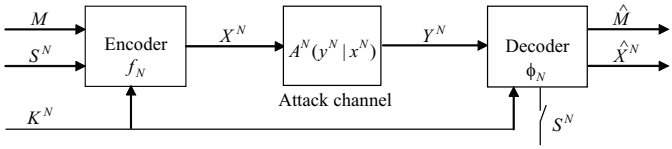

Figure 1: Information communication problem

M is the hiding information, $S^N = (S_1, S_2, \cdots, S_N)$ is host data set, $K^N = (K_1, K_2, \cdots, K_N)$ is side information. M is to be hidden into S^N through the side information K^N. Encoder must select a good method to implement this hiding work, i.e. to find a good expression for $f_N = (M, S^N, K^N)$. After encoding, we get embedded information X^N, which is certain to suffer from some attack in information communication, then the decoder receives Y^N from the communication channel $A^N(y^N \mid x^N)$, the decoder must try his best to get the hidden information through $\phi_N(Y^N, K^N)$. The message $\hat{M}(\hat{M} \in \wp)$ must have some difference with M due to the attack.

To the encoder, he must guarantee the transparency and robustness of information hiding, so X^N should have a little difference with S^N, in order to control the difference; the encoding progress should be subject to the expected-distortion constraint as Equation (1):

$$Ed^N(S^N, X^N) \leq D_1. \tag{1}$$

Meanwhile, there is a relative formula as follows:

$$d^N(x^N, y^N) = \frac{1}{N}\sum_{k=1}^{N} d(x_k, y_k).$$

To the attacker, he must guarantee his attack is acceptable, otherwise Y^N has nothing to do with X^N. In that way, his attack loses meaning, so the attack progress should be subject to some distortion constraint.

$$Ed^N(X^N, Y^N) \le D_2. \quad (2)$$

Then $C(D_1, D_2)$ is defined as the super-mum of all achievable rates for distortions D_1, D_2. A rate $R = \frac{1}{N}\log|\wp|$ is achievable for distortions (D_1, D_2), too. Literature [5] presents a more concrete expression for Equation (1), which assumes that $\wp$ follows proportional distribution, the dependencies between S and K are modelled by a joint distribution $p(s,k)$, then $\wp$ is independent of S and K, so Equation (1) can be expressed as:

$$\sum_{s^N\in S^N}\sum_{k^N\in K^N}\sum_{m\in\wp}\frac{1}{|\wp|}p(s^N,k^N)d_1^N(s^N, f_N(s^N,m,k^N)) \le D_1.$$

Equation (2) can be expressed as:

$$\sum_{x^N\in X^N}\sum_{y^N\in Y^N} d_2^N(x^N,y^N)A(y^N|x^N)p(x^N) \le D_2.$$

The average probability of error is described as:

$$\begin{aligned} P_{e,N} &= \frac{1}{|\wp|}\sum_{m\in\wp} Pr[\phi_N(Y^N,K^N)\neq m|M=m] \\ &= \frac{1}{|\wp|}\sum_{m}\sum_{s^N}\sum_{(y^N,k^N):\phi N(y^N,k^N)\neq m} \\ &\qquad A^N(y^N|f^N(S^N,m,k^N))p(s^N,k^N). \end{aligned}$$

In the recent years, many models have been set up, but they are all subject to analogy distortion constraint. For example, M. Barni et al have put forward their ideas in [1]. They consider that information hiding consists in the modification of a set of DCT, DFT, or DWT coefficients, and the amount of modification each coefficient undergoes is proportional to the magnitude of the coefficient itself as expressed by the following rule:

$$x_i = s_i + \gamma m_i|s_i|. \quad (3)$$

Where s_i indicates the original DCT, DFT, or DWT coefficients, x_i indicates the hided coefficients, m_i is the i-th component of the hiding information, i presents the position of the marked coefficients within the frequency spectrum, and γ is a parameter controlling the hiding strength.

Cover and Chiang have designed four channel particularities in [3], the four channels denote four special case of channels:

1) Channel C_{00}: neither the encoder nor the decoder knows the side information.

2) Channel C_{11}: both the encoder and the decoder know the side information.

3) Channel C_{01}: only the decoder knows the side information.

4) Channel C_{10}: only the encoder knows the side information.

Then Cover and Chiang research on hiding capacity of the four channels according to the character of them.

3 Encode, Attack and Decode

In order to guarantee the transparency and robustness of information hiding, encoding must be constrained by the exception-distortion (1). At the same time, encoder should consider the reasonability and complexity of his algorithm.

As we can see that many encoders would like to hide their information into image, video, audio and so on. When they hide information into image, they consider that embedding information into the frequency of the image is safer than just embedding into the original one. They may select discrete cosine transform, discrete Fourier transform, discrete wavelet transform or Walsh transform. Moulin and O'Sullivan have compared their hiding capacity [6], especially, they compare Block-DCT with Wavelet EQ model and the comparison reveals that capacity estimating under the Wavelet EQ model is lower than that under the Block-DCT model. Since both expressions are upper bounds on actual capacity, so we should consider more concrete distribution of information when we choose transformation.

Before they decide to select a kind of transformation they must consider its characters first. After that they should decide which coefficient can be modified without corrupting transparency. Mostly they choose the coefficients that have large variance around them. Barni et al. [1] have analyzed the changeful trend of image data hiding capacity as the coefficient variance increases, and explained the reason of the result.

Barni et al. hide information according to their rule, which is expressed in Equation (3), this kind of algorithm is widely used. However, if encoder uses this algorithm, the decoder must have the host information S^N, otherwise, he cannot extract the hide information. In many cases, decoder has not the host data set, so the encoder needs other algorithms to deal with. A great deal of algorithms which can realize blind examination are manipulated by adjusting the linear connection between coefficients, so the decoder can get hidden information by analyzing the relation in coefficients, a good example is used in [10].

From Figure 1 we can get $Y^N = A^N(y^N|x^N)$. An attacker passes X^N through a random attack channel $A^N(y^N|x^N)$ to produce corrupted data Y^N. Here

$A^N(y^N|x^N)$ depicts the statistics dependence between input signals and output ones. The decoder who has plentiful experiences can deduce which kind of attack has taken place, and then he may extract hidden information more exactly.

The capacity estimates of data hiding systems under some practical attacks must be considered because the attack is inevitable. Information hiding can be thought as a game among two cooperative players (the information hider and the decoder) and an opponent (the attacker). The first player tries to maximize a payoff function, the opponent tries his best to minimize it, so the function should have connection with encode f_N, decode ϕ_N, attack A^N. Surely, the game must obey the two expected-distortion constraints (1) and (2), so the hiding-capacity is the function of $f_N, \phi_N, A^N, D_1, D_2$, we define payoff function as follows:

$$\begin{aligned} & J(f_N, \phi_N, A^N, D_1, D_2) \\ = & \max_{f_N, D_1, \phi_N} \{ \min_{A_N, D_2} \{C(f_N, \phi_N, A^N, D_1, D_2)\}\}. \end{aligned}$$

If D_1, D_2 is decided at first, then

$$J(f_N, \phi_N, A^N) = J(f_N, \phi_N, A^N, D_1, D_2).$$

Here, let's define the support set of $p(s,k)$, $\Omega = \{(s,k) \in S \times K : p(s,k) > 0\}$.

We introduce an auxiliary variable U (Maybe U includes some information about host data, side information, attack channel, or nothing important). We use $Q(x,u|s,k)$ denote a conditional probability distribution function (pdf) from $S \times K$ to $X \times U$. The function $Q(x,u|s,k)$ has depicted all information about encoding procession. Furthermore, it includes some other information with respect to U. Additionally, ϕ_N is determined by f_N in a certain degree, so $Q(x,u|s,k)$ can substitute f_N and ϕ_N, we can derive that:

$$J(Q, A) = J(f_N, \phi_N, A^N).$$

Then we can get data hiding - capacity by the application of information theory:

$$J(Q, A) = I(U; Y|k) - I(U; S|K).$$

This equation denotes that payoff function is equal to the discrimination between the quantity of information that can be got about U from Y and the quantity of information that can be got about U from S under the same assumption that side information is known.

Conclusion 1: Assume that for any $N \geq 1$, the attacker knows f_N, the decoder knows both f_N and the attack channel, a rate R is achievable for distortion D_1 and attacks in the class $\{A(f_N)\}$, if and only if $R < C$, where $C = \max_{Q(x,u|s,k)\in Q} \min_{A(y|x)\in A(Q)} J(Q, A)$ and U is a random variable defined over an alphabet U of cardinality of $|U| \leq |X||\Omega| + 1$.

4 Estimates of Data Hiding - Capacity

Recent research has shown that the data hiding capacity is the value of a mutual-information game among the encoder, decoder and attacker. The capacity of channel is given as $C = \max_{p(M)}\{I(M; Y^N)\}$ by Claude Shannon in [8]. In this section, we will give some conclusions about the capacity of some kinds of channels and models, such as non-blind channels, blind channels, gauss channels, parallel gauss channels, and AR-1 models as well.

4.1 Non-Blind Channels

4.1.1 A Simple Non-Blind Case

Let's first discuss a simple non-blind case, the data hiding capacity can be written as follows:

$$\begin{aligned} C &= \max_{p(M)}\{I(M, Y^N|S^N)\} && (4) \\ &= \max_{p(M)}\{H(M, S^N) - H(M|S^N, Y^N)\} && (5) \\ &\leq H(M_*|S^N). && (6) \end{aligned}$$

Equation (4) is the definition of the capacity for the non-blind case (decoder knows host data set S^N). Equation (5) is the definition of the mutual-information by the theory of information. Equation (6) is reached because of the non-negativeness of entropy.

4.1.2 Gauss Channels

If S follows a Gaussian distribution, let $S \sim N(0, \delta^2)$, $d(x,y) = (x-y)^2$, this model is very special and widely used. Literature [2, 5] gives an explicit solution to the estimation of upper bounds on capacity of non-blind Gaussian distribution S. They reached the following conclusion:

Conclusion 2: Let $S = X = Y = R$ (R is the set of real number) and $d(x,y) = (x-y)^2$ be the squared-error distortion measure. Assume that $K = S$, let α be the maximized of the following function:

$$f(\alpha) = \frac{[(2\alpha-1)\delta^2 - D_2 + D_1][D_1 - (\alpha-1)^2\delta^2]}{[D_1 + (2\alpha-1)\delta^2]D_2},$$

in the interval $(\alpha_{\inf}, 1 + \sqrt{D_1}/\delta)$, where $\alpha_{\inf} = \max(1, \frac{\delta^2+D_2-D_1}{2\delta^2})$. Then we have:

1) If $D_2 \geq (\delta + \sqrt{D_1})^2$, the hiding capacity is $C = 0$.

2) If S is non-Gaussian with mean zero and standard deviation $\delta > \sqrt{D_2} - \sqrt{D_1}$, the hiding capacity is upper-bounded by

$$\begin{aligned} C_G = \frac{1}{2}\log(1 + \\ \frac{[(2\alpha-1)\delta^2 - D_2 + D_1][D_1 - (\alpha-1)^2\delta^2]}{[D_1 + (2\alpha-1)\delta^2]D_2}). \end{aligned} \quad (7)$$

3) If $S \sim N(0, \delta^2)$ and $D_2 < (\delta + \sqrt{D_1})^2$, the hiding capacity is given by Equation (7). The optimal covert channel is given by $X = \alpha S + Z$, where $Z \sim N(0, D_1 - (\alpha - 1)^2\delta^2)$ is independent of S. The optimal attack is the Gaussian test channel from rate-distortion theory,

$$A^*(y|x) = N(\beta^{-1}x, \beta^{-1}D_2), \tag{8}$$

where $\beta = \frac{\delta_x^2}{\delta_x^2 - D_2}$, and $\delta_x^2 = D_1 + (2\alpha - 1)\delta^2$.

The role of the host-signal scaling parameter $\alpha \geq 1$ in Conclusion 2 is to increase the value of δ_x^2 and thereby to reduce the effective noise variance δ_w^2 of the Gaussian test channel. From Equation (7) we can give $C = \frac{1}{2}\log(1 + \frac{\delta_z^2}{\delta_w^2})$, where δ_w^2 decreases as α increases, and δ_z^2 increases as tends to 1. Hence the optimal value of α results from a tradeoff.

4.1.3 Parallel Gauss Channels

Parallel Gaussian models are useful in that they are reasonably tractable and provide capacity expressions for realistic signal models. They also provide upper bounds on capacity if the actual distribution of S differs from the model are a more important reason. For instance, any correlation between subsignals S_k would decrease capacity, as well as any deviation from a Gaussian distribution with the same second-order statistics [7]. For non-Gaussian S, the Gaussian upper bound on capacity is asymptotically tight as D_1 and D_2 approach zero, and in this case the capacity-achieving distributions are the same as in the parallel-Gaussian case. A fundamental implication of this result is that the exact distribution of the source plays only a second-order effect in a small-distortion scenario.

4.1.4 Spike Models

Let us consider a rate-distortion bound for still-image compression, under a so-called spike model that captures the sparsity of wavelet image representations [9]. It appears that this model is very useful in data hiding as well. Under a spike model, there are two types of channels: those with large variance $\delta_k^2 << D_1, D_2$ and those with low variance $\delta_k^2 >> D_1, D_2$. The signal components are independent. Assume that $\delta_k^2 >> D_1, D_2$ for $1 \leq k \leq K^*$ and $\delta_k^2 << D_1, D_2$ for $K^* < k < K$. Then, we will get $C = \frac{1}{2}r^* \log(1 + \frac{D_1}{D_2 - D_1})$, where $r^* = \sum_{k=1}^{K^*} r_k \in [0, 1]$ is the fraction of strong signal component. In conclusion, for spike models:

1) The optimal power allocations by the data hider and the attacker are independent of the signal variances in the strong channels, provided that these variances are much relative to D_1 and D_2.

2) The optimal data-hiding strategy equalizes the power among the strong channels, and likewise, the optimal attack strategy equalizes the noise power among strong channels. Negligible power is allocated to weak channels.

3) The (per-sample) capacity C_k is the same for all strong channels and is negligible for the weak channels. The capacities $\{C_k\}$ are in the strong channels and $C = \sum_k r_k C_k$ depends only on the distortion levels D_1 and D_2, rather than the variances δ_k^2.

4.1.5 AR-1 Models

AR-1 model is a classical model. First, we assume the host-image source is a separable AR-1 Gaussian process [4]. The distribution of this process is parameterized by four quantities: mean μ, variance δ^2, horizontal and vertical correlation coefficients ρ_x and ρ_y respectively. The two-dimensional (2-D) spectral density of S is given by

$$S(f_x, f_y) = \frac{\delta^2(1 - \rho_x^2)(1 - \rho_y^2)}{|1 - \rho_x e^{-j2\pi f_x}|^2 |1 - \rho_y e^{-j2\pi f_y}|^2},$$

$$-\frac{1}{2} \leq f_x, f_y \leq \frac{1}{2}.$$

Compute data-hiding capacity estimates for image sources characterized by different values of ρ_x, ρ_y and δ^2. The capacities are computed using the numerical algorithm mentioned in [6]. Finally, several results are reached:

1) There is a threshold of 2 D_2 to C to be satisfactory in the sense that the approximation is quite accurate (except for very low values of $D_2/D_1 - 1$).

2) There is a saturation of the capacity C_k in a given channel when the variance δ_k^2 in that channel increases, and C_k is proportional to δ_k^2 for small δ_k^2.

3) For relatively low values of $\rho_x = \rho_y$ (say less than 0.8), capacity is essentially the same as in the i.i.d. Gaussian case: $C \approx \frac{1}{2}\log(1 + \frac{D_1}{D_2 - D_1})$.

4) As $\rho_x = \rho_y$ approaches 1, capacity tends to zero, as more and more channels (frequencies) become weak and hence are unable to hide significant information.

4.2 Blind Channels

Rate of reliable transmission for blind information hiding clearly cannot be higher than the rate in the case that the decoder has access to side information and host data. So hiding-capacity is upper-bounded by Equation (7) for any $p(s)$. In the following, Theorem 5.3 in [5] has given the optimal blind-information-hiding strategy and optimal attack for Gaussian $p(s)$. The optimal attack $A(y|x)$ comes from the Gaussian test channel (8).

Conclusions show that the hiding-capacity for blind channels is the same as Equation (7), i.e. the achievable rate of reliable transmission is the same whether or not the host data are known of the decoder. Of course, Equation (7) is an upper bound on hiding capacity if S is non-Gaussian with mean zero and variance δ^2.

Now, we continue to analyze the channels in [3]. Cover and Chiang have assumed that the channel is embedded in some environment with state information S_1^N available to the sender, correlated state information S_2^N available to the receiver, and a memoryless channel with transition probability $p(y|x, s_1, s_2)$ that depends on the input X and the state S_1, S_2 of the environment.

We assume that $(S_{1,i}, S_{2,i})$ are i.i.d. $p(s_1, s_2)$, $i = 1, 2 \cdots$. The output has conditional distribution $p(y^N|s_1^N, s_2^N) = \prod_{i=1}^N p(y_i|x_i, s_{1,i}, s_{2,i})$.

Conclusion 3: The memoryless channel $p(y|x, s_1, s_2)$ with state information $(S_{1,i}, S_{2,i})$ i.i.d. $\sim (s_1, s_2)$, with S_1^N available to the sender and S_2^N available to the receiver, has capacity of

$$C = \max_{p(u,x|s_1)} [I(U; S_2, Y) - I(U; S_1)].$$

Conclusion 4: The four capacities with state information are special cases of Conclusion 3: $C_{00} = \max_{p(x)} I(X; Y)$, $C_{11} = \max_{p(x|s)} I(X; Y|S)$, $C_{01} = \max_{p(x)} I(X; Y|S)$, $C_{10} = \max_{p(u,x|s)} [I(U; Y) - I(U; S)]$.

Conclusion 5: For a bounded distortion measure $d(x, \hat{x})$ and $(X_i, S_{1,i}, S_{2,i})$ i.i.d. $\sim p(x, s_1, s_2)$, where X, S_1, S_2 are finite sets. Let S_1^N be available to the encoder and S_2^N to the decoder. The rate distortion function is $R(D_1, D_2) = \min_{p(u|x,s_1)p(\hat{x}|u,s_2)} [I(U; S_1, X) - I(U; S_2)]$, where the minimization is under the distortion constraint (e1) and (e3).

Conclusion 6: The four rate distortion functions with state information are special cases of Conclusion 5: $R_{00} = \min_{p(\hat{x}|x)} I(X; \hat{X})$, $R_{11} = \min_{p(\hat{x}|x,s)} I(X; \hat{X}|S)$, $R_{10} = \min_{p(\hat{x}|x)} I(X; \hat{X})$, $R_{01} = \min_{p(u|x)p(\hat{x}|u,s)} [I(U; X) - I(U; S)]$.

5 Conclusions and Expectations

Many researchers have characterized data-hiding capacity for realistic image sources, or other host data, by the application of information theory on some kinds of channels. This paper has made a summary of the work about data-hiding capacity, several statistical image models have been considered. We hope that those results can be a guard when deciding how much bits of secret information can be hidden into the host data without corrupting the transparency and robustness.

References

[1] M. Barni, R. Bartolini, A. De Rosa, and Apia. "Capacity of watermarked-channel: How many bits can be hidden within a digital image?" *Proceedings of SPIE*, vol. 3657, pp. 437-448, Jan. 1999.

[2] T. M. Cover and J. A. Thomas. "Elements of Information Theory", New York: Wiley, 1991.

[3] T. M. Cover and M. Chiang. "Duality between channel capacity and rate distortion with two-sided state information", *IEEE Transactions on Information Theory*, vol. 48, no. 6, June 2002.

[4] A. K. Jain. "Fundamentals of digital image processing", Englewood Cliffs, NJ: Prentice-Hall, 1989.

[5] P. Moulin and J. A. O'Sullivan. "Information-theoretic analysis of information hiding", in *IEEE International Symposia on Information Theory*, pp. 17-34, Boston, MA, Aug. 1988.

[6] P. Moulin and J. A. O'Sullivan. "A framework for evaluating the data-hiding capacity of image sources", *IEEE Transaction on Image Processing*, vol. 11, no. 9, pp. 124-135, Sept. 2002.

[7] P. Moulin and M. K. Mihcak. "The parallel-Gaussian watermarking game", UIUC Coord. Sci. Lab Tech. Report, CS514, Jan. 2002.

[8] C. E. Shannon. "A mathematical theory of communication", *BELLSYS*, vol. 27, pp.379-423, 1948.

[9] C. Weidmann and M. Vetterli. "Rate-distortion analysis of spike processes", in *Proceedings of the Data Compression Conference*, pp. 45-54, Snowbird, UT, Mar. 1999.

[10] Y. ZHU and L. Xia. "An algorithm for blind digital watermark based on DCT coefficient", *Mini Micro Computer System*, vol. 24, no. 3, pp. 571-573, Mar. 2003.

Zhensong Liao male, born in 1979, a Ph.D. candidate. He studies in the department of computer science and technology, in Huazhong University of Science and Technology. He majors in Grid Security and Grid Performance Analysis.

Yan Huang female, born in 1980, a Ph.D. candidate. She is currently working toward the Ph.D. degree in the Department of Control Science and Engineering, Huazhong University of Science and Technology. Her research interests include image and video processing, compression, information hiding, complex dynamics and computer assistant proof.

Chisong Li female, born in 1976, a Ph.D. candidate and an assistant. She studies in the department of computer science and technology, in Huazhong University of Science and Technology. She majors in Grid Security and High Performance Computing.