Journal of Information Hiding and Multimedia Signal Processing ©2012 ISSN 2073-4212
Ubiquitous International Volume **3**, Number **2**, April **2012**

The Capacity of a Stegosystem for the Noisy Attack Channel

Valery Korzhik
State University of Telecommunication
61, Moika, 191186, St. Petersburg, Russia
val-korzhik@yandex.ru

Guillermo Morales-Luna
Computer Science Department
CINVESTAV-IPN
Av. IPN 2508, 07360, Mexico City, Mexico
gmorales@cs.cinvestav.mx

Ksenia Nebaeva
State University of Telecommunication
61, Moika, 191186, St. Petersburg, Russia
ksenya_2002@mail.ru

Received October 2011; revised April 2012

Abstract. *We consider a scenario where an attacker is able to receive a stegosignal only over a noisy channel under the condition that he (or she) may even know the cover message (CM). We propose to use a spread-spectrum-signal-based stegosystem (SG) and prove that the capacity of such SG is zero for any fixed security level. Moreover, we show that the code rate satisfies a square root law due to Ker* et al. *(in reality this law can be found in an earlier paper (2005) by the authors of the current paper). We get the same result for a conventional SG where the CM is modeled by a Gaussian correlated sequence with a correlation matrix unknown for the SG designer. We speculate that for any conventional SG with noiseless attack channel but for a practical CM model, the capacity should also be zero.*

Keywords: Watermarking, stegosystem, noisy channel, blind decoder.

1. **Introduction.** The notion of *channel capacity* has been introduced by the pioneering Shannon's work [1] at 1948. It is determined as the maximum (or supremum in general) information that can be transmitted over noisy channel. There are different types of noisy channels for which the channel capacity was calculated in a closed form depending on the channel parameters: symmetric channel without memory, z-channel, Gaussian band limited channel, fading channel, etc [2]. Later, more "exotic" channels were investigated in order to find their capacity. The most outstanding result has been achieved by A. Wyner [3] in 1975 for the concept of wire-tap channel capacity that is determined as the maximal information that can be transmitted over the channel under the condition of zero-information leaking for the wiretapper.

The most essential significance of this notion is due to Shannon's theorem asserting: *if the code rate R is lower than the capacity C then there exist coding and decoding algorithms providing a decreasing to zero error probability as the code block length approaches to infinity.* This theorem belongs to the class of the so called *existence theorems* where encoding and decoding methods are not specified. In a post Shannon period many thousands of papers were devoted to finding such methods (more generally modulation and demodulation) which are approaching the Shannon bound.

There are several mathematical tools to prove an exponential decreasing of the error probability as the block length approaches to infinity, and Shannon's condition holds. But in the case of a noisy channel model with unlimited frequency bandwidth this proof can be given in a very simple form. We present now this proof because it will be requested in the next Section of the current paper.

Let us consider $m = 2^k$ mutually orthogonal continuous signals, each of length T (it is possible because, due to our assumption, our channel is frequency unlimited).

Let us use an additive bound [4] in order to get the following formula for the error probability of an m-ary orthogonal signaling with coherent receiver [5]:

$$P_e \leq (2^k - 1)Q(h) \tag{1}$$

where

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{+\infty} e^{-\frac{t^2}{2}}\,dt \qquad , \qquad h = \sqrt{\frac{P_S T}{N_0}},$$

P_S is the power of orthogonal (indeed orthonormal) signals, and N_0 is the spectral power density of white Gaussian channel noise. Using a well known bound for Q, namely

$$Q(x) \leq \frac{1}{2} e^{-\frac{x^2}{2}}$$

and substituting this inequality into (1) we get after a simple transform:

$$\begin{aligned} P_e &\leq \frac{1}{2} 2^k \exp\left(-\frac{P_S T}{2N_0}\right) \\ &= \frac{1}{2} 2^{\nu T} \exp\left(-\frac{P_S T}{2N_0}\right) \\ &= \frac{1}{2} e^{(\ln 2)\nu T} \exp\left(-\frac{P_S T}{2N_0}\right) \\ &= \frac{1}{2} \exp\left(-T\left(\frac{P_S}{2N_0} - \nu \ln 2\right)\right) \end{aligned} \tag{2}$$

where $\nu = \frac{k}{T}$ is the code rate in bits/second.

We can see that $P_e \to 0$ under the condition that N_0, P_S, ν remain constant as $T \to +\infty$ and the following condition holds

$$\nu < \frac{P_S}{(2 \ln 2)N_0} = 0.72 \frac{P_S}{N_0}. \tag{3}$$

The right side of (3) fixes a lower bound for the channel capacity because more precise calculations show that in reality the channel capacity is $1.44 \frac{P_S}{N_0}$ (see [2]).

Although the real channel capacity is at least twice as the bound (3), it is unlikely that if the additive bound-based capacity is zero, then the real capacity is not zero. We will use this claim in the following Section.

There are very well known papers devoted to the investigation on stegosystems capacity [17, 11, 12, 16] but, as a rule, they execute with a description of the CM as a Gaussian i.i.d., or as a correlated Gaussian sequence with known correlation matrices, or even the CM is supposed given as a Markov chain with known transition matrices. All these approaches are highly impractical. Linguistic SG and Network SG are outside of our consideration. It is indeed quite strange to find the capacity for the ε-security SG [17] where it is determined as normalized relative entropy (or KLD) calculated just through one sample. In reality this notion does not determine a reasonable SG security connected with a pair of probabilities P_m and P_{fa} which depend on the blocks of length n.

In order to be successful within the risk of unknown CM statistics, we apply the concept of noisy channel introduced in [6]. Then the attackers problem consists in a statistical distinguishing of the CM after its passing over the noisy channel and the SG signal passing over the same noisy channel. Since the channel noise distribution is, as a rule, known much better (mostly it is a Gaussian one) the problem to investigate such a SG is simplified.

2. **The stegosystem capacity for a noisy attack channel.** Let us consider the concept of a stegosystem (SG) *based on a noisy attack channel* proposed in [6].

This setting can be justified only if there exists in a natural manner a noisy channel and the attacker is able to receive the stegosignal just over this channel, and nothing else. There are a lot of examples where such a situation takes place. For the thing: slightly noised musical files which are placed on some sites in the Internet and are free for reading, audio and video signals in the Skype system where noises are present due to bad cameras and microphones, speech signals within cell phone communications, etc.

Let us consider a binary linear block code V with parameters (n, k) where n is the block length and k is the number of information bits. The embedding of a secret message is performed in an additive manner as follows:

$$C_w(\ell) = C(\ell) + (-1)^{b_{i\ell}} \sigma_w \pi(\ell) \ , \qquad \ell = 1, 2 \ldots, n,$$

where $C_w(\ell)$ is the ℓ-th sample of the stegosignal, $C(\ell)$ is the ℓ-th sample of the cover message (CM), $b_{i\ell}$ is the ℓ-th bit of the i-th codeword of length n, $\sigma_w \geq 0$ is the coefficient that determines the depth of the embedding, and π is a zero-mean Gaussian pseudorandom i.i.d. sequence with variance 1 determined by the stegokey.

After a passing of the stegosignal through the Gaussian channel we get

$$C'_w(\ell) = C_w(\ell) + \varepsilon(\ell) \ , \qquad \ell = 1, 2 \ldots, n,$$

where C'_w is the signal observed by an attacker, and ε is a zero-mean Gaussian i.i.d. sequence with variance σ_ε^2.

We assume also (in favour of the attacker) that he (or she) may even know the CM, which is unacceptable for conventional SG.

The informed legal decoder takes a decision about the embedding of the i-th codeword by making

$$i = \arg \max_{1 \leq i' \leq 2^k} \sum_{\ell=1}^{n} \left(C'_w(\ell) - C(\ell)\right) (-1)^{b_{i'\ell}} \pi(\ell). \tag{4}$$

The attacker's problem consists in statistically distinguishing the channel noise within the sum of the channel noise and the embedded SG signal. Since the channel noise and the stegosignal are both Gaussian zero-mean i.i.d. sequences, the attacker must perform a testing of two hypothesis H_0 and H_1 regarding the variances of the sequences σ_ε^2 or $\sigma_\varepsilon^2 + \sigma_w^2$, respectively.

It is a very simple problem in mathematical statistics, but for our purposes it is more convenient to use an information-theoretic measure of hypothesis testing, namely the *relative entropy* $\mathbf{D}(H_0 \| H_1)$ [7, 8]. It follows from Information Theory [8] that for any hypothesis testing rule, the following inequality holds

$$P_m \ln\left(\frac{P_m}{1 - P_{fa}}\right) + (1 - P_m) \ln\left(\frac{1 - P_m}{P_{fa}}\right) \leq \mathbf{D}(H_0 \| H_1), \tag{5}$$

where P_{fa} is the *probability of false alarm* (when the SG signal has not been embedded but the detector wrongly declares its presence) and P_m is the *probability of missing* (when the SG signal has been embedded but the detector wrongly declares its absence). The SG system is *perfect* or *completely secure* if $\mathbf{D}(H_0\|H_1) = 0$ and then $P_{fa} = P_m = \frac{1}{2}$ which is equivalent to a random guessing about the SG presence or the SG absence. If $\mathbf{D}(H_0\|H_1) \neq 0$ but has an enoughly small value, then we may assume that the SG is *D-secure.* If we select a proper distinguishing threshold such that $P_m = 0$, then it follows from (5) that $P_{fa} \geq 2^{-D}$.

It has been proved in [9] that for a chosen model of two zero mean i.i.d. Gaussian distributions with different variances σ_ε^2 and $\sigma_\varepsilon^2 + \sigma_w^2$ we get

$$\mathbf{D}(H_1\|H_0) = 0.77\, n \left[\ln\left(1+\eta\right)^{-1} - \left(1+\eta\right)^{-1}\right], \tag{6}$$

where

$$\eta = \frac{\sigma_\varepsilon^2}{\sigma_w^2}$$

is the *noise-to-watermark ratio* (NWR). For typically large NWR η we may approximate (6) as

$$\mathbf{D}(H_0\|H_1) = 0.36\, n\eta^{-2}. \tag{7}$$

We get immediately from (7) that

$$\eta = 0.6\sqrt{\frac{n}{D}}. \tag{8}$$

Using the informed decoder rule (4) and the additive bound for the code V with minimal code distance d we get similar to (1) and using the equation 16 from [6]:

$$P_e \leq (2^k - 1)\, Q\left(\sqrt{\frac{d}{\eta}}\right).$$

Following the transformation technique used in (2), we get the inequality

$$P_e \leq \exp\left(Rn\ln 2 - \frac{d}{2\eta}\right), \tag{9}$$

where $R = \frac{k}{n}$ is the code rate. Substituting η from (8) into (9) and taking into account the fact that for any (n, k)-block code, $d \leq n$, we get from (9)

$$\begin{aligned} P_e &\leq \exp\left(Rn\ln 2 - \frac{n\sqrt{D}}{1.2\sqrt{n}}\right) \\ &= \exp\left(Rn\ln 2 - \frac{1}{1.2}\sqrt{nD}\right) \end{aligned} \tag{10}$$

We can see from (10) that $P_e \to 0$ as $n \to +\infty$ for any security level $D > 0$ only if

$$R \leq 1.2\sqrt{\frac{D}{n}}. \tag{11}$$

The inequality (11) means that the capacity of the SG system for a noisy attack channel for any given security level D is zero because there does not exist a parameter $C > 0$ for which $P_e \to 0$ as $n \to +\infty$ for $R < C$. (We define the *capacity of the SG system* as the maximum code rate for which $P_e \to 0$ as $n \to +\infty$ for any (not necessarily perfect) security level D.)

This result is quite reasonable and it coincides with a claim asserted in [10] that the greater is the block length the greater is the information that can be used by an attacker in order to distinguish the CM and the SG system.

We can find also from (11) that the code rate obey to square root laws of steganographic capacity [11] due to Ker [12]. (But in reality, a similar square root law has been established for the noisy attack channel in an earlier paper [9]). It follows from (8) that in order to provide a given constant security level D as $n \to +\infty$, the NWR should decrease to zero and this results in an impossibility of practical implementation of such SG system using digital CM and watermarking. In order to overskip this unfortunate situation, the *spread-time stegosystem* (STS) has been proposed in [13], where the embedding of a pseudorandom sequence $(\pi(\ell))_{\ell=1}^{n}$ is performed not in every CM sample but in time spreaded samples which are determined by a secret stegokey. But as it was shown in [13] it is necessary to embed $\pi(n)$ only in at most $\sqrt{n}$ samples among the block length n in order to provide the conditions estableshing constant η and $\mathbf{D}(H_0\|H_1)$. Hence a square root law also holds. It is worth to note that no Dirty Paper Codes [14] can make the SG capacity to be non-zero for this case, because we assume that an informed decoder is used, hence these codes are not needed. In fact a bottleneck to our approach appears only in the use of an additive bound for a proof of zero capacity but, as it was mentioned in Section 1, it is very unlikely that we get a non-zero capacity using some better bound (say, similar to Gallager exponential bound [2]).

3. **Some remarks regarding capacity of conventional stegosystems.** It is easy to get a non-zero capacity of the stegosystem for artificial model of a CM, say Gaussian zero-mean i.i.d. sequence with known variance σ_c^2. Using the modified additive embedding rule (so far for the uncoded case)

$$C_w(\ell) = \alpha C(\ell) + (-1)^b \sigma_w \pi(\ell) \ , \qquad \ell = 1, 2 \dots, n, \tag{12}$$

where $\alpha = 1 - \frac{B}{\sigma_c^2}$, $\sigma_w^2 = B\left(1 - \frac{B}{4\sigma_c^2}\right)$, $b \in \{0, 1\}$, B is the CM quadratic distortion after embedding, and π is a zero-mean Gaussian i.i.d. pseudorandom sequence with variance 1 determined by a stegokey, we get a perfect SG with $\mathbf{D}(H_0\|H_1) = 0$ and then the capacity of such SG, with an attack by additive noise and even blind decoder, is non-zero. But such CM model is very far from real CM.

Let us consider the closer to practice case where the CM is also Gaussian zero-mean but a correlated sequence determined by its correlation matrix R_{c^n} for the n-block of samples. If the designer of the SG would know R_{c^n}, then he (or she) would be able to perform the *Karhunen-Loève Transform* (KLT) that reduces the correlated sequence to an i.i.d. sequence and to get thus an ideal SG, after the embedding of the secret message b using (12) and after performing the inverse KLT [16]. But this assumption is also very far from practice, moreover the correlation matrix R_{c^n} may differ for different sample n-blocks.

Therefore we may assume that R_{c^n} is unknown for the SG designer that embeds the message according to rule (12). In the paper [15], the Bhattacharyya coefficient (connected directly with the *Bhattacharyya distance*) was proposed as a new criterion for the SG security because sometimes it is more convenient than the relative entropy. Many years ago, the Bhattacharyya distance has been effectively used by T. Kailath [18] as a technique to estimate the probability of error for optimal receiver under the condition of binary signaling over noisy channel

For any SG hypothesis testing rule, the Bayesian error probability $P = \frac{1}{2}\left(P_m + P_{fa}\right)$ (for equal prior probability of the SG presence and absence) satisfies the inequalities [15]:

$$\frac{1}{4}\rho^2 \leq P \leq \frac{1}{2}\rho. \tag{13}$$

For such model it has been proved in [15] that for the particular case of an exponential correlation matrix R_{c^n}, the square Bhattacharyya coefficient ρ can be expressed as

$$\rho^2 \approx (A(r, \eta_w))^{n-1} \tag{14}$$

where $A(r, \eta_w)$ is some coefficient that depends on the parameter r of the correlation matrix R_{c^n} and on the *signal-to-watermark ratio* (SWR) with embedding distortion B

$$\eta_w = \frac{\sigma_c^2}{B}.$$

We can see from (14) that in order to keep a constant SG security level determined by the probability P in the left side of (13), we should increase the SWR as the block length n is increasing. Since σ_c^2 is kept constant, this means that *the distortion B should be given by a decreasing function of n.* It has been proved in [15] that the probability of error for an uncoded SG with embedding given by (12) and the informed decoding rule

$$\Lambda = \sum_{\ell=1}^{n} \left(C'_w(\ell) - \alpha C(\ell)\right) \pi(\ell) \ \Rightarrow \ \tilde{b} = \begin{cases} 1 & \text{if } \Lambda \geq 0 \\ 0 & \text{if } \Lambda < 0 \end{cases}$$

where $C'_w(\ell) = C_w(\ell) + \varepsilon(\ell)$, $\varepsilon(\ell) = N(0, \sigma_\varepsilon^2)$, $\varepsilon(\ell) \in$ i.i.d., can be presented for large η_w, as follows

$$P_e = Q\left(\sqrt{\frac{n}{\eta - 1}}\right),$$

where

$$\eta = \frac{\eta_w}{\eta_a} \quad , \quad \eta_a = \frac{\sigma_c^2}{B + \sigma_c^2}.$$

Then using an additive bound and the bound for the function Q, we get, similar to (9)

$$P_e \leq \exp\left(Rn \ln 2 - \frac{d}{2(\eta - 1)}\right).$$

Assuming as before that $d \leq n$, for any block code we get the following inequality for the code rate R providing an exponential decreasing of the block error probability P_e as the block length n approaches to infinity:

$$R \leq 0.72\,(\eta - 1)^{-1}. \tag{15}$$

Because the noise power σ_ε^2 is constant and (as we showed before) the SNR η_w is an increasing function with n, for small B we get

$$\eta - 1 \approx \eta = \frac{\sigma_c^2(B + \sigma_\varepsilon^2)}{B\sigma_c^2} = \frac{\sigma_\varepsilon^2}{B}. \tag{16}$$

We can see from (16) that $(\eta - 1)^{-1}$ is a monotonically decreasing function of n and therefore it follows from (15) that the capacity of such SG is zero.

Moreover, with “the point of view of an additive bound”, any SG for which in order to keep constant the security level with any increasing of the block length, it is required to decrease the WNR, and hence it has zero capacity. This looks as speculation, but indeed such property seems to be valid for any real digital SG. By the way, this fact should not completely discourage SG designers because even in this case it is possible to embed into CM some amount of secure bits but only at the cost to increase the length of the CM.

4. **Conclusions.** The capacity of the watermarking systems (without restriction on their detectability) is a very reasonable notion, but the capacity of stegosystems, which is determined for a given level of their security (undetectability) except for ideal (undetectable) SG, is a very questionable notion for real SG. The notion "real SG" can be understood as SG based on CM statistics close to real (practical) situations.

In the current paper we focus to show that the capacity of SG for noisy attack channel is zero. This is our main contribution. Our proof is based not on a strong upper bound for the SG capacity but on the lower union bound for the probability of block error. But if we change an additive bound to a more precise Gallager exponent bound [2], then it results in the same conclusion. Of course it gives also a lower bound for the capacity, but for conventional communication channels (say discrete symmetric memory-less channels) this bound coincides with real capacity.

We showed also that for a CM modeled by Gaussian zero-mean correlated sequences, with correlation matrix unknown for the SG designer, the SG capacity is also zero. It seems to be very likely that for any SG for which we have to decrease WNR with the increasing of block length in order to keep constant a security level, the capacity is zero. This claim is very close to the *Anderson's concept* as formulated in [10].

REFERENCES

[1] C. E. Shannon, A mathematical theory of communication, *Bell System Technical Journal*, vol. 27, pp. 379-423, pp. 623-656, 1948.

[2] R. G. Gallager, *Information Theory and Reliable Communication*, USA, Wiley, 1968.

[3] A. Wyner, Wire-tap channel concept, *Bell System Technical Journal*, vol. 54, pp. 1355-1387, 1975.

[4] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*, USA, John Wiley & Sons, 1965.

[5] J. Proakis, *Digital Communications, Fourth Edition*, USA, McGraw-Hill, 2001.

[6] V. Korjik, M. H. Lee, and G. M. Luna, Stegosystems based on noisy channels, *Proc. of the 9th Spanish Meeting on Cryptology and Information Security*, pp. 379-387, 2006.

[7] C. Cachin, An information-theoretic model for steganography, *Proc. of 2nd International Workshop on Information Hiding*, pp. 306-318, 1998.

[8] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, USA, John Wiley, 1991.

[9] V. Korzhik, G. M. Luna, and M. H. Lee, On the existence of perfect stegosystems, emphProc. of 4th International Workshop on Information Hiding, pp. 30-38, 2005.

[10] R. Anderson and F. Petitcolas, On the limits of steganography, *IEEE Journal of Selected Areas in Communications*, vol. 16, no. 4, pp. 474-481, 1998.

[11] L. Fearnley, Square root laws of steganographic capacity in the context of existing models, 2009, `http://www.cs.auckland.ac.nz/courses/compsci725s2c/archive/termpapers/lfearnley.pdf`.

[12] A. D. Ker, T. Pevný, J. Kodovský, and J. Fridrich, The square root law of steganographic capacity, *Proc. of the 10th ACM workshop on Multimedia and security*, pp. 107-116, 2008.

[13] V. Korzhik, G. M. Luna, K. Loban, and I. M. Begoc, Undetectable spread-time stegosystem based on noisy channels, *Proc. of the International Multiconference on Computer Science and Information Technology*, pp. 723-728, 2010.

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

[15] V. I. Korzhik, H. Imai, J. Shikata, G. Morales-Luna, and E. Gerling, On the use of Bhattacharyya distance as a measure of the detectability of steganographic systems, *Trans. Data Hiding and Multimedia Security*, vol. 3, pp. 23-32, 2008.

[16] Y. Wang and P. Moulin, Steganalysis of block-structured stegotext, *Proceedings of SPIE on Security, Steganography, and Watermarking of Multimedia Contents*, vol. 5306, pp. 477-488, 2004.

[17] P. Comesaña and F. Pérez-González, On the capacity of stegosystems, *Proc. of the 9th Workshop on Multimedia & Security*, pp. 15-24, 2007.

[18] T. Kailath, The Divergence and Bhattacharyya Distance Measures in Signal Selection, *IEEE Trans. Communications*, vol. 15, no.1, pp. 52-60, 1967.