Секция «Теория вероятностей и математическая статистика»

On optimization and application of Ryabko & Ryabko asymptotically optimal perfect steganographic scheme

Научный руководитель – Kabatiansky Grigory

Kaziakhmedov E.A.[1], *Melnikov G.A.*[2], *Kireev K.S.*[3]

1 - Сколковский институт науки и технологий, Информационные технологии, Moscow, Russia, *E-mail: edgar.kaziakhmedov@skoltech.ru*; 2 - Сколковский институт науки и технологий, Информационные технологии, Moscow, Russia, *E-mail: grmuller1996@gmail.com*; 3 - Сколковский институт науки и технологий, Информационные технологии, Moscow, Russia, *E-mail: klim.s.kireev@gmail.com*

Аннотация.
We study in terms of efficiency an asymptotically optimal perfect steganographic system[1] described in [2]. We propose an algorithm of selection optimal parameters for encoding and decoding for specific message. An application of this method to digital images is also considered.

Steganography is a technique which allows to hide secret data in an ordinary message (also called container or cover message), resulting message is usually called stegoword. The main aim of a steganographic algorithm is to prevent the fact of hiding from being revealed. Since a detection method which is used by an observer is usually unknown, universal criteria for estimation of stegosystem are required.

There are two main approaches to address this problem. One of them came from practice and can be called combinatiorial. The idea is to measure the number of changes between the stegoword and the cover message, i.e. their hamming distance. For reliable communication this distance should be minimized. This criterion emerged in work of Crandell[4] where the author described the method called matrix encoding. This method can be briefly explained as usage of linear covering codes to reduce number of changes, so it can be studied via coding theory[5]. The minimal number of changes per encoded bit of information may appear to be the most practical parameter for optimization, nevertheless there are stegoanalisys algorithms which allow to detect embedding with high probability[]. It occurs, that the positions of modified bits and types of modification are also important due to possible changes in cover message distribution [].

To take into account this embedding property, the second criterion was proposed. In the approach described by Cachin[6], it is assumed that the cover message is generated by a source with its particular probability distribution, which should not be modified by embedding. In [6] the stegosystem with this property is called *perfect*. The construction of perfect stegosystems for sources with independent symbols was proposed and studied in [1]. In this paper we investigate embedding efficiency of this system. Our main results are estimation of embedding efficiency, proposing the algorithm of optimizing parameter selection and implementation this system on digital images.

1. RYABKO SYSTEM

1.1. General Description

In this part, we generally describe the embedding algorithm proposed in [1]. Although, in original paper an arbitrary alphabet was considered in this report we decided to study binary alphabet, because real data usually can be approximated as binary sequence.

Let μ be source of i.i.d. random variables. Let u be a word with length n generated with μ consisted of 0s and 1s. The secret message is τ. Let us denote $S(u)$ as set of all permutations of u sorted in certain order, *e.g.* lexicographical. Consider the binary representation of $|S(u)|$:

$$|S(u)| = (a_{m-1}, ..., a_1, a_0)$$

where $m = [log_2(|S(u)|)]$, let us denote $\delta(u)$ as order number of u in $S(u)$ and consider binary representation of $\delta(u)$:

$$\delta(u) = (\lambda_{m-1}, ..., \lambda_1, \lambda_0)$$

Let j be $max(i : a_i \neq \lambda_i)$. We can embed j bits of secret message τ to u via choosing stegoword v with following rule:

$$\delta(v) = (\lambda_{m-1}, ..., \lambda_{j+1}, \tau_j, ..., \tau_0)$$

$S(u) = S(v)$, therefore $j_u = j_v$ and the message must be completely reconstructed on decoding side. In [1] it was proved that u and v are independent random valuables, so distribution must not be affected by this operation.

1.2. Example

For a better explanation let us involve an example:
Encoding:

$$u = 1010, \tau = 11_2$$

$$S(u) = \begin{bmatrix} 1100 \\ 1010 \\ 1001 \\ 0110 \\ 0101 \\ 0011 \end{bmatrix}$$

$$\delta(u) = 1 = 0001_2$$

$$|S(u)| = 6 = 0110_2$$

$$j = 2$$

$$\delta(v) = 0011_2 = 3 \implies v = 0110$$

Decoding:

$$v = 0110_2$$

$$\delta(v) = 3$$

$$j = 2$$

$$\tau = 11_2$$

2. EFFICIENCY AND CAPACITY

In [] two metrics of estimation stegosystem were proposed, one of them is *capacity* denoted as $C(u)$ in present report is the number of embedded bits per one bit of u, another one is *efficiency* denoted as $E(u, \tau)$ which is the number of embedded bits per one modified bit (in the word u):

$$E(u,\tau) = \frac{C(u)n}{D(u,\tau)}$$

where $D(u,\tau)$ is a hamming distance between u and v. In the example above we encoded 2 bits to the 4-bit word, so $C(u) = 2, d(u,v) = 2$, therefore $E(u,\tau) = 1$ Since the secret massage is random and has uniform distribution, $E(u,\tau)$ can be averaged to obtain $E(u)$.

2.1. Efficiency

Теорема 1. *If u is a cover message with length n containing k 1s, $k \leq \frac{n}{2}$, and τ is a random secret message with uniform distribution.*

$$|S(u)| = (a_{m-1}, ..., a_1, a_0)$$

Then:

$$E(u) = \frac{\sum_{i=0}^{m} a_i i 2^i}{\sum_{j=0}^{k} 2i\binom{k}{j}\binom{n-k}{j}}$$

Доказательство. In [1] it was proved that:

$$C(u)n = L(u) = \frac{1}{|S(u)|}\sum_{i=0}^{m} a_i i 2^i$$

To prove bottom expression, it suffices to find a sum of hamming distances between u and all elements of $S(u)$. Any permutation of u with hamming distance $2i$ can be treated as a selection of i ones and i zeroes. A number of ways to do it is $\binom{k}{j}\binom{n-k}{j}$. Iterating over all i:

$$D(u) = \frac{1}{|S(u)|}\sum_{j=0}^{k} 2i\binom{k}{j}\binom{n-k}{j}$$

$$E(u) = \frac{L(u)n}{D(u)}$$

Then:

$$E(u) = \frac{\sum_{i=0}^{m} a_i i 2^i}{\sum_{j=0}^{k} 2i\binom{k}{j}\binom{n-k}{j}}$$

3. APPLICATION

To test this stegosystem we implemented it for .png images. Emdedding is performed to last significant bit. We count long-run occurrences to show differences in distribution between the stegoword and the cover message.

Список литературы

[1] Ryabko B. Ya., Ryabko D.B "Asymptotically optimal perfect steganographic systems Prob. Peredach. Inform., vol. 45, no. 2, pp.119-126, 2009.

[2] Steganalysis, High Capacity Despite Better, and Andreas Westfeld. "F5?A Steganographic Algorithm."In Information Hiding: 4th International Workshop, IH 2001, Pittsburgh, PA, USA, April 25-27, 2001. Proceedings, vol. 2137, p. 289. Springer Science & Business Media, 2001.

[3] Simmons G. J., "The Prisoner's Problem and the Subliminal Channel Advances in Cryptology, Proceedings of CRYPTO' 83 (Workshop on Communication Security), Plenum, New York, pp. 51-67, 1984.

[4] Crandall R.,"Some notes on steganography", 1998

[5] F. Galand, G. Kabatiansky, "Steganography via covering codes", Proceedings. IEEE International Symposium on Information Theory, p.192, 2003.

[6] Cachin C., "An Information-Theoretic Model for Steganography Proc. 2nd Int. Workshop on Information Hiding, Lecture Notes Comput. Sci., 1525, Springer, Berlin, pp. 306-318, 1998.

Рис. 1. Picture before embedding

Рис. 2. Picture after embedding

Рис. 3. Long-run occurences test