STEGANOGRAPHIC METHOD WITH HISTOGRAM CORRECTION

E.O. VOLKOREZ
Belarusian State University
Minsk, BELARUS
e-mail: evgenijvolkorez@yandex.ru

Abstract

Efficient steganographic method with histogram correction is introduced. Method implies solution of optimization problem to minimize both histogram and hamming distortion of container. Fast solution of the problem is shown.

1 Introduction

The purpose of steganography is to make communication undetectable. It is achieved by embedding data in container by slightly modifying them. The most important property of any steganographic method is statistical undetectability. There are two known approaches in steganography. Covering codes [1] provides low number of embedding changes. Statistical restoration [2] eliminates histogram distortion it is restored by modifying symbols from reserved set. Statistical restoration can be combined with almost any embedding method (including covering codes) at the cost of lower throughput and some additional distortion.

A method, presented in the paper, provides both low number of embedding changes and eliminates histogram distortion. It is generalization and improvement of method [3]. It has the next advantages compared to statistical restoration combined with covering codes:

- method has the same throughput as covering codes,
- method implies less additional distortion per 1 data bit embedded.

The main drawback of presented method is its complexity, as it requires to solve an optimisation problem. However problem is shown to have a fast solution.

2 Embedding method

Let data set D and container set C be a linear spaces over the binary finite field $\mathbb{F}_2$:

$$D = \mathbb{F}_2^m, \quad C = \mathbb{F}_2^n.$$

Definition 1. Linear embedding method is the pair of functions (F, ϕ_F), where $F : C \to D$ is a linear surjective function, $\phi : C \times D \to C$ satisfies equation $\phi(c, d) \in F^{-1}(d)$. We'll denote F as an extraction function, ϕ_F as an embedding function.

Definition 2. Let $\rho : C \times C \to \mathbb{R}$ be a distance function. An embedding method is optimal, if it provides minimal distortion:

$$\phi(c, d) \in \arg\min_{\tilde{c} \in F^{-1}(d)} \rho(c, \tilde{c}). \tag{1}$$

Container histogram distortion can be easily detected, so embedding method has to leave histogram unchanged. In addition method should modify as low container elements as possible to avoid detection by analysis of other features. Therefore let consider two container distortion measures:

- $w(c_1 - c_0)$ number of changed container elements,
- $|w(c_1) - w(c_0)|$ container weight change, or histogram distortion,

where w is a number of nonzero vector elements. Define distance function ρ as follows:

$$\rho(c_0, c_1) = W|w(c_1) - w(c_0)| + w(c_1 - c_0), \quad W > w(c) \quad \forall c \in C.$$

Due to the definition of W, the optimal embedding function can be determined by the optimization problem with two criteria:

$$\begin{array}{l} |w(\phi) - w(c)| \to min, \\ |w(\phi - c)| \to min, \\ \phi \in f^{-1}(d). \end{array} \tag{2}$$

Lemma 1. *Embedding function has next representation:*

$$\phi = c + e_0 + k, \quad k \in F^{-1}(0),$$

where $e_0 = \arg\min_{c' \in F^{-1}(d - F(c))} w(c')$.

Introduce following characteristics of container modification by value $k \in F^{-1}(0)$:

- $h = w(c + e_0 + k) - w(c + e_0)$ correction value,
- $w = w(e_0 + k) - w(e_0)$ correction distortion,
- $u = \frac{w}{h}$ correction specific distortion.

Theorem 1. *Problem (2) is equivalent to the following problem:*

$$\begin{array}{l} |h + h_0| \to \min, \\ hu \to \min, \\ (h, u) \in P, \end{array} \tag{3}$$

where P is the set of allowable parameters:

$$P \subset \left\{(h(c, d), k(c, d)), u(c, d, k) | k \in f^{-1}(0)\right\}.$$

Remark 1. Let $h_0 = w(c + e_0) - w(c)$ be the histogram distortion of container modification by e_0. We further assume that $h_0 \leq 0$, the case $h_0 \geq 0$ is very similar.

Consider the ordering relationship on the set of all possible parameters:

$$(h_1, u_1) < (h_2, u_2) \Leftrightarrow \begin{cases} h_1 \leq h_2, \\ u_1 > u_2. \end{cases}$$

Remark 2. If the inequality meets $(h_1, u_1) < (h_2, u_2)$, then (h_1, u_1) can't be a solution of problem (3). Therefore remove such parameters from set P:

$$(h_1, u_1) \in P \Rightarrow \nexists (h, u) \in P : (h_1, u_1) < (h, u).$$

Sort elements of the set P in the ascending order with the primary key h and the secondary key u, the result is the parameters sequence $p = (p_i)_{i=1}^{I}$, where $I = |P|$.

3 Block embedding method

Definition 3. (F, ϕ) is a block method, if following equations are hold:

$$\begin{array}{ll} D = D^1 \times ... \times D^L, & F(c) = (F^1(c^1), F^2(c^2), ..., F^L(c^L)), \\ C = C^1 \times ... \times C^L, & \phi(c, d) = (\phi^1(c^1, d^1), \phi^2(c^2, d^2), ..., \phi^L(c^L, d^L)), \end{array}$$

where (F^l, ϕ^l) linear embedding method defined on data set D^l containers set C^l. Set D^l, C^l a subspaces of linear spaces D and C.

Previously defined notations of embedding method in relation to method (F^l, ϕ^l) we will denote with upper index l, e.g. p^l is the parameters sequence.

Theorem 2. *For the block embedding method the optimisation problem (3) is equivalent to the following problem:*

$$\begin{array}{l} |\sum h^l + h| \to \min, \\ \sum_{l=1}^{L} h^l u^l \to \min, \\ (h^l, u^l) \in P^l. \end{array} \qquad (4)$$

Consider two cases:

1. Inequality (5) holds. The solution of the problem (4) is the sequence $(p^l_{I^l})_{l=1}^{L}$.

$$\sum_{l=1,..,L} p^l_{I^l} < -h_0, \qquad (5)$$

2. Inequality (5) doesn't hold. Solution reduced to the auxiliary problem (6).

Theorem 3. *If the inequality (5) doesn't hold, problem (4) is equivalent to the auxiliary problem (6):*

$$\begin{array}{l} \sum_{l=1}^{L} \sum_{j=1}^{i^l} \Delta u_j^l \Delta h_j^l \to \min, \\ \sum_{l=1}^{L} \sum_{j=1}^{i^l} \Delta h_j^l \geq -h, \\ i^l \in [1, I^l] \in \mathbb{Z}, \end{array} \qquad (6)$$

Relation between problems (4) and (6) is given by equations (7),(8):

$$\Delta p_1 = (0, u_1^l), \quad \Delta p_l = \left(h_i^{l+1} - h_i^l, \tfrac{w_{i+1}^l - w_i^l}{h_{i+1}^l - h_i^l}\right). \tag{7}$$

$$h_i^l = \textstyle\sum_{j=1}^i \Delta h_j^l, \qquad u_i^l h_i^l = \textstyle\sum_{j=1}^i \Delta u_j^l \Delta h_i^l. \tag{8}$$

Consider $\tilde{p}$ the subsequence of p:

$$\tilde{p}_i^l = p_{j_i}^l, i = 1, ..., \tilde{I}^l.$$

Indexes j_i are defined according to iterative formula (9), where values $\Delta\tilde{p}^l$ are given by applying equation (7) to sequence $\tilde{p}^l$.

$$\begin{array}{ll} j_{\tilde{I}^l} & = I^l, \\ j_{\tilde{I}^l - 1} & = I^l - 1, \end{array} \quad j_i^l = \max\left\{ j \,\middle|\, j < j_{i+1}^l, \frac{w_{j_i+1}^l - w_j^l}{h_{j_i+1}^l - h_j^l} < \Delta\tilde{u}_{j_i+1} \right\}. \tag{9}$$

Join elements of sequences $(\Delta\tilde{p}^l)_{l=1}^L$ into single sequence s. Sort s in the ascending order with the primary key u and the secondary key h:

$$s = (s_i)_{i=1}^I = (h_s^i, u_s^i)_{i=1}^I, \quad I = \sum_{l=1}^L \tilde{I}^l.$$

Let b_i^l be the number of i's element of sequence $\Delta\tilde{p}^l$ in the sequence s.

Theorem 4 (The solution of problem (6)). *If equality (10) meets for some $i_0 \in \mathbb{Z}$:*

$$-h = \sum_{i=1}^{i_0} h_s^i, \tag{10}$$

then the solution of problem (6) is given by the following formula:

$$i^l = j_{\substack{\max(i), \\ b_i^l <= i_0}}^l \quad l = 1, ..., L. \tag{11}$$

Remark 3. Application of theorem 4 is limited by condition (10), but in practice strict correction isn't required. Instead of h_0 we can take the nearest value $|h_0'| < |h_0|$, that meets condition (10). It means that (10) can be changed to formula (12):

$$i_0 = \max_{\substack{\sum_{j=1}^i h_s^j \leq -h_0, \\ i=1, ..., I.}} (i)\,. \tag{12}$$

References

[1] Bierbrauer J., Fridrich J.(2008) Constructing good covering codes for applications in steganography. *LNCS*. Vol. **4920**, pp. 1-22.

[2] Solanski K. et al.(2005) Statistical restoration for robust and secure steganography, *ICIP*. Vol. **II**, pp. 1118-1121.

[3] Volkorez E.O. (2009) Steganographic algorithm, making use of the binary Hamming codes for data embedding and distortion correction. *IST2009*. Vol. **2**, pp. 24-25.