СТЕГАНОГРАФИЧЕСКИЙ АЛГОРИТМ, ИСПОЛЬЗУЮЩИЙ БИНАРНЫЕ КОДЫ ХЕММИНГА ДЛЯ ВСТРАИВАНИЯ ДАННЫХ И КОРРЕКТИРОВКИ ИСКАЖЕНИЙ

Е.О. Волкорез

НИИ прикладных проблем математики и информатики,
НИЛ статистического анализа и моделирования
пр. Независимости 4, г. Минск, РБ
телефон: 8(029) 9337663; e-mail: veoveo@tut.by

Предложен алгоритм встраивания данных в бинарную последовательность, использующий бинарные коды Хэмминга, как для встраивания данных, так и для корректировки возникающих искажений статистических характеристик контейнера. Исследованы статистические характеристики алгоритма.

Ключевые слова – коды Хэмминга, корректировка, стеганография.

1 ВВЕДЕНИЕ

Задача скрытой передачи данных посредством мультимедиа контейнера, часто сводится к встраиванию данных в бинарную последовательность. Например, в последовательность младших бит квантованных коэффициентов дискретного косинусного преобразования JPEG изображения. Проблема эффективного встраивания данных в последовательность впервые была рассмотрена в статье [1] и в дальнейшем была сведена к конструированию кодов в работах [2,3].

Основное условие скрытой передачи данных – стойкость к обнаружению факта передачи данных. На данный момент, наиболее действенные методы обнаружения скрытых данных базируются на анализе гистограммы контейнера, а также на анализе более сложных статистических характеристик.

В статье предложен метод, основанный на применении свойств кодов Хэмминга, и обладающий двумя основными преимуществами:

- большой объем встроенных данных при малой доле искажений контейнера,
- корректировка возникающих при встраивании статистических характеристик, за счет внесения относительно небольшого объема искажений.

2 АЛГОРИТМ ВСТРАИВАНИЯ ДАННЫХ

Встраивание данных происходит путем модификации бинарной последовательности. Встраивание происходит поблочно: последовательность разбивается на L блоков длины n, в каждый отдельный блок встраивается вектор данных длины m. В каждый блок данные встраиваются с помощью кодов Хэмминга согласно одному из нескольких вариантов. Выбор подходящей модификации каждого блока позволяет минимизировать заранее выбранную меру искажений контейнера. Другими словами, алгоритм встраивания данных состоит из двух более простых:

1. алгоритм встраивания данных в блок, обеспечивающий несколько вариантов модификации,
2. алгоритм выбора оптимальной модификации каждого из блоков.

Рассмотрим алгоритм встраивания данных $d \in B^m$ в блок $b \in B^n$, где $n = 2^m - 1$, $m \in Z^+$. Определим синдром вектора b по формуле:

$$s = s(b) = \sum_{i=1}^{n} B(i) b_i , \qquad (1)$$

где $B(i)$ – бинарное представление числа i. Пусть $Z^+(a)$ - представление бинарного вектора a целым положительным числом, $\delta_{i,j}$ – символ Кронекера.

Встраивание данных – поиск такого вектора $\tilde{b}$, что $s(\tilde{b}) = d$. Стандартный алгоритм встраивания изменяет максимум один элемент b согласно формуле:

$$\tilde{b} = b - \left\{ \delta_{i,Z^+(s-d)} \right\}_{i=1}^{n} .$$

Корректирующий алгоритм допускает изменение двух элементов вектора b, что позволяет выбрать лучший из нескольких вариантов его модификации $j_0 \in [0,n]$:

$$\tilde{b} = b - e^{s,d,j_0} , \qquad (2)$$

$$e^{s,d,j} = \left\{ \delta_{i,j} + \delta_{i,Z^+(s-d-B(j))} \right\}_{i=1}^{n} . \qquad (3)$$

Утверждение 1. Алгоритм (1-3) корректен, т.е. для любого $j_0 \in [0,n]$ верно равенство:

$$s(\tilde{b}) = d .$$

Также стандартный алгоритм является частным случаем корректирующего при $j = 0$ и $j = Z^+(s-d)$. Кроме того, если $s \neq d$, то существует ровно $\frac{n+1}{2} = 2^{m-1}$ различных вариантов модификации вектора b. Если $s = d$, т.е. все варианты модификации совпадают и $\tilde{b} = b$.

Алгоритм выбора оптимальной модификации блоков контейнера, минимизирующий искажений гистограммы, состоит из следующих шагов:

1. Оценка параметров блоков:

a. изменения числа 0 при встраивании данных стандартным алгоритмом (1-3):

$$\Delta b^l = 2\,|\,b_{Z^+(s-d)}\,| - 1,$$

где $|\,b_0\,| = 0.5$;

b. характеристика максимального $\Delta b^l_{\max}$ и минимального $\Delta b^l_{\min}$ увеличения числа нулей в блоке:

$$\Delta b^l_{\max} = 2 \max_{j=0,n}\left(\left|b_j\right| + \left|b_{Z^+(s-d-B(j))}\right|\right) - 2 - \Delta b^l,$$

$$\Delta b^l_{\min} = 2 \min_{j=0,n}\left(\left|b_j\right| + \left|b_{Z^+(s-d-B(j))}\right|\right) - 2 - \Delta b^l.$$

2. Оценка суммарного объема искажений Δx при использовании стандартного алгоритма:

$$\Delta x = \sum_{l=1}^{L} \Delta b^l.$$

3. Вычисление параметров корректировки:

a. Вычисление эффективности корректировки блока – модуль величины, на которую корректируется частота нулей в рамках блока:

$$eff_l = \begin{cases} \Delta b^l - \Delta b^l_{\min}, \text{если } \Delta x \ge 0, \\ \Delta b^l_{\max} - \Delta b^l, \text{иначе;} \end{cases}$$

b. Оценка числа блоков, имеющих эффективность корректировки $eff = 0,1,3$:

$$n(eff) = \sum_{l=1}^{L} \delta_{eff,\,eff_l};$$

c. Оценка числа блоков, к которым будет применен корректирующий алгоритм:

$$N(3) = \begin{cases} n(3), & \text{если } 3n(3) < |\,\Delta x\,|, \\ \left[\Delta x / 3\right], & \text{иначе,} \end{cases}$$

$$N(1) = \begin{cases} n(1), & \text{если } n(1) < \Delta x - 3n(3), \\ \left[\Delta x - 3n(3)\right], & \text{иначе,} \end{cases}$$

$$N(0) = 0.$$

4. Встраивание данных. Если $\Delta x < 0$, то данные встраиваются корректирующим алгоритмом в блок исходя из максимизации числа нулей, иначе исходя из минимизации. Корректирующий алгоритм используется для встраивания данных в $N(eff)$ случайно выбранных блоков, среди имеющих эффективность eff. В остальные блоки данные встраиваются согласно стандартному алгоритму.

4 ИССЛЕДОВАНИЕ СТАТИСТИЧЕСКИХ ХАРАКТЕРИСТИК АЛГОРИТМА

Исследуем вероятность корректировки уменьшения частот 0 предложенным алгоритмом в случае, когда последовательность данных и контейнера являются последовательностями испытаний Бернулли с вероятностями единицы равными 0.5 и p_1 соответственно.

Утверждение 2. Если данные встраиваются во все блоки согласно стандартному алгоритму, то верна асимптотика:

$$\frac{\Delta x}{L} \underset{L\to\infty}{\sim} N\left(\mu_0, L^{-1}(1-\mu_0)\mu_0\right),$$

где $\mu_0 = p_{s\neq d}(p_1 - p_0)$.

Утверждение 3. Матожидание μ и второй момент μ_2 максимального увеличения числа нулей блока $\Lambda b^l_{\max}$ определяются следующим образом:

$$\mu = p_{s\neq d}\left(2 - (2-p_1)P_2 - p_0^n\right),$$

$$\mu_2 = p_{s\neq d}\left(4 - (4-p_1)P_2 + p_0^n\right),$$

где $P_2 = (1-p_1^2)^{(n-1)/2}$, $p_{s\neq d} = P(s \neq d)$.

Утверждение 4. Если в каждый блок данные встраиваются исходя из максимизации числа 0, то есть $n(3) = N(3)$, $n(1) = N(1)$, то изменение числа нулей $\Delta x_{\max}$ распределено асимптотически нормально:

$$\frac{\Delta x_{\max}}{L} \underset{L\to\infty}{\sim} N(\mu, L^{-1}\sigma^2),$$

где $\sigma^2 = \mu_2 - \mu^2$.

Следствие 1. Вероятность полной корректировки уменьшения числа 0, возникающего в результате встраивания данных, определяется следующим образом:

$$P(\Delta x_{\max} \ge 0) \underset{L\to\infty}{\cong} F\left(L^{-0.5}\sigma^{-1}\mu\right), \qquad (4)$$

где F – функция нормального распределения.

Следствие 2. Для сходимости вероятности полной корректировки (4) к 1 необходимо выполнение неравенства $\mu > 0$. Для $m = 2,\ldots,8$, решения указанного неравенства, имеет вид:

$$p_1 > p_1^*(m),$$

$$\{p_1^*(m)\}_{m=2}^{8} \approx \{0.27, 0.16, 0.097, 0.058, 0.034, 0.02, 0.011\}.$$

ЛИТЕРАТУРА

[1] Crandall, R. Some notes on steganography / R. Crandall // 1998.

[2] Bierbrauer, J. Constructing good covering codes for applications in steganography / J. Bierbrauer, J. Fridrich // LNCS 4920 – Berlin ,2008 – pp. 1–22.

[3] Zhang, W. Steganographic codes – a new problem in coding theory / W. Zhang, S. Li // 2002.