МОДИФИКАЦИЯ СТЕГАНОГРАФИЧЕСКОГО АЛГОРИТМА НА ОСНОВЕ КОДОВ ХЭММИНГА

А. М. КАПУСТА

При встраивании информации в изображения формата JPEG встраивание осуществляется в область преобразования, т. е. в коэффициенты дискретного косинусного преобразования (ДКП). При этом обычно допускают изменение только наименее значимых битов (НЗБ) ненулевых квантованных ДКП-коэффициентов, чтобы вносимые при этом искажения в изображение были минимальны.

Задачу встраивания информации в изображение можно свести к задаче встраивания информации в бинарную последовательность. Необходимо произвести такую модификацию исходной бинарной последовательности $x = (x_1, \ldots, x_N) \in B^N$ длины N, что из модифицированной последовательности $\tilde{x} = (\tilde{x}_1, \ldots, \tilde{x}_N) \in B^N$ может быть однозначно извлечена встроенная информация $y = (y_1, \ldots, y_M) \in B^M$ длины M, где $B = \{0, 1\}$.

При встраивании информации даже при отсутствии визуальных искажений могут быть обнаружены искажения статистических характеристик контейнера, в первую очередь гистограмм ДКП-коэффициентов [3]. Следовательно, чтобы повысить стойкость стеганографического метода, необходимо уменьшить искажения контейнера, что вызовет снижение пропускной способности. И наоборот, для повышения пропускной способности придется пожертвовать стойкостью. Соответственно, актуальной задачей является конструирование методов, обеспечивающих наилучшее практическое соотношение стойкости и пропускной способности.

Проблема эффективного встраивания данных впервые была рассмотрена Крэндэллом [1]. Он показал, что использование линейных кодов может существенно улучшить соотношение объема встроенных данных и числа искаженных бит контейнера. Идея Крэндэлла была развита Бирбрауэром, проблема эффективного встраивания данных была сведена к конструированию стеганографических кодов [1]. Данный подход с использованием кодов Хэмминга впервые был применен в стеганоалгоритме F5 [6], где он получил название «матричное кодирование».

Другое направление исследований в области обеспечения стеганографической стойкости основывается на применении алгоритмов с корректировкой искажений статистических характеристик, возникших при встраивании информации. Применяемые алгоритмы двухшаговые: на первом шаге часть НЗБ ДКП-коэффициентов используется непосредственно для встраивания информации, на втором шаге оставшаяся часть НЗБ ДКП-коэффициентов используется для корректировки искажений статистических характеристик, возникших на первом шаге. Такой подход затруднит использование статистических методов стеганоанализа.

Впервые данный подход был применен в алгоритме OutGuess [4], где проводилась корректировка искажений глобальной гистограммы ДКП-коэффициентов. В более совершенном алгоритме MB2 [5] корректируются искажения блочности, глобальной и локальных гистограмм.

В данной работе предложена модификация стеганографического алгоритма на основе кодов Хэмминга с целью повышения стойкости к статистическому стеганоанализу.

Специфика формата JPEG состоит в том, что изображение разбивается на блоки пикселей размера 8×8, внутри каждого блока производится двумерное ДКП, а полученные ДКП-коэффициенты квантуются согласно таблице квантования, которая зависит от степени сжатия изображения. Обозначим квантованные ДКП-коэффициенты c_{ij}^k, где $i, j = 1, \ldots, 8$ определяют позицию коэффициента внутри блока, а $k = 1, \ldots, B$ определяет номер блока, B – общее число блоков 8×8 ДКП-коэффициентов в изображении.

Определим глобальную гистограмму $H = (H_L, \dots H_R)$ всех $8 \times 8 \times B$ ДКП-коэффициентов:

$$H_s = \sum_{i=1}^{8}\sum_{j=1}^{8}\sum_{k=1}^{B} I\{s = c_{ij}^k\},\ L \le s \le R,$$

где $L = \min\limits_{ijk} c_j^k$, $R = \max\limits_{ijk} c_j^k$, $I(A) = 1$, если верно A, иначе $I(A) = 0$.

Аналогичным образом определим локальную гистограмму $H^{ij} = (H_L^{ij}, \dots, H_R^{ij})$ для каждой позиции (i, j) ДКП-коэффициента:

$$H_s^{ij} = \sum_{k=1}^{B} I\{s = c_{ij}^k\},\ L \le s \le R,\ 1 \le i, j \le 8.$$

Рассмотрим алгоритм встраивания двоичного вектора $d \in B^m$ в вектор $b \in B^n$, где $n = 2^m - 1$, $m \in \mathrm{N}$, методом синдромного кодирования на основе кодов Хэмминга.

Определим синдром вектора b по следующей формуле:

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

где $B(i)$ – бинарное представление числа i в виде битовой строки из множества B^n.

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

$$\tilde{b} = b - e_{s-d},$$

где вектор $e_0 = 0$, а вектор e_j, $j = 1, \dots, n-1$, в бинарном представлении имеет только одну единицу на позиции j:

$$e_{j,k} = \delta_{j,k},\ 1 \le k \le n,$$

здесь $\delta_{i,j}$ – символ Кронекера.

Предложим модифицированный алгоритм, который допускает изменение не более чем двух элементов вектора b, что позволяет выбрать лучший из нескольких вариантов модификации:

$$\tilde{b} = b - e_{s-s'} - e_{s'-d},\ 0 \le s' \le n-1,$$

где s' определяет наилучший вариант модификации.

Корректность модифицированного алгоритма следует из равенств:

$$s(e_{s-d}) = s - d,$$

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

Если $s \ne d$, то стандартный алгоритм изменит соотношение числа единиц и нулей при встраивании информации, так как изменится один из битов вектора b. В модифицированном алгоритме будет доступно $\frac{n+1}{2}$ вариантов модификации вектора b, из которых можно выбрать наилучший вариант, когда не изменяется соотношение числа единиц и нулей при встраивании информации.

Представим модифицированный ***алгоритм встраивания информации:***

1. Ненулевые квантованные ДКП-коэффициенты перемешиваются в случайном порядке на основе генератора случайных чисел, инициализируемом значением секретного ключа.

2. Для каждого вида ДКП-коэффициента, определяемого позицией *(i, j)*, $i, j = 1,...,8$, в блоке 8×8, и тех их значений, которые различаются только наименее значимым битом, строятся независимые бинарные подпоследовательности.

3. Каждая бинарная подпоследовательность разбивается на блоки длины $n = 2^m - 1$ согласно параметру кода Хэмминга m.

4. В блок длины n встраивается m бит информации методом синдромного кодирования, при этом исказиться может не более 2 бит блока.

5. Восстанавливается исходный порядок ДКП-коэффициентов, при этом часть из них будет иметь инвертированный младший бит вследствие встраивания информации.

Для проведения экспериментов была использована выборка из 1000 изображений в формате JPEG. Встраиваемая информация представляет собой случайный бинарный вектор с равной вероятностью появления нулей и единиц $p\{0\} = p\{1\} = 0,5$.

Введем меру отклонения гистограммы контейнера после встраивания информации, используя следующую формулу:

$$D_h = \sqrt{\sum_{i=L}^{R} (H_i - \widetilde{H}_i)^2} ,$$

где L и R – минимальное и максимальное значения квантованных ДКП-коэффициентов; H_i – доля ДКП-коэффициентов со значением i в исходном изображении, а $\widetilde{H}_i$ – в модифицированном контейнере.

В таблице 1 представлены результаты сравнения отклонения гистограмм при встраивании информации стандартным и модифицированным алгоритмами на основе кодов Хэмминга при различных значениях параметра кода m – в $2^m - 1$ бит контейнера встраивается m бит сообщения.

Таблица 1

Искажение гистограмм ДКП-коэффициентов при встраивании

Параметры кода	Стандартный алгоритм			Модифицированный алгоритм		
	Откл. глоб. гист.	Макс. откл. лок. гист.	Ср. откл. лок. гист.	Откл. глоб. гист.	Макс. откл. лок. гист.	Ср. откл. лок. гист.
2	0,905	2,650	1,271	0,517	1,662	0,735
3	0,445	1,388	0,686	0,080	0,391	0,142
4	0,218	0,750	0,386	0,005	0,068	0,010
5	0,105	0,423	0,228	0,001	0,010	0,000
6	0,051	0,250	0,140	0,000	0,002	0,000
7	0,024	0,152	0,088	0,000	0,001	0,000

Как показывают проведенные эксперименты, модифицированный алгоритм при той же пропускной способности позволяет добиться гораздо меньшего отклонения гистограммы ДКП-коэффициентов, чем стандартный алгоритм встраивания на основе кодов Хэмминга, следовательно, он обеспечивает лучшую стойкость к методам статистического стеганоанализа.

Список использованных источников

1. *Bierbrauer, J.* Constructing good covering codes for applications in steganography. / J. Bierbrauer, J. Fridrich // Lecture Notes in Computer Science. – 2006. – Vol. 4920. – P. 1–22.

2. *Crandall, R.* Some notes on steganography / R. Crandall // Steganography Mailing List. – 1998.

3. *Fridrich , J.* Steganalysis of JPEG images: breaking the F5 algorithm / J. Fridrich, M. Goljan, D. Hogea // Information hiding, 5th international workshop. – N. Y., 2003. – P.~310–323.

4. *Provos, N.* Defending against statistical steganalysis / N. Provos // Proc. of the 10 USENIX Security Simposium. – 2001. – P.~323–325.

5. *Sallee, P.* Model-based methods for steganography and steganalysis / P. Sallee // International Journal of Image and Graphics. – 2005. – Vol.~5~(1). – P. 167–190.

6. *Westfeld, A. A.* High capacity despite better steganalysis (F5 – a steganographic algorithm) / A. A. Westfeld // Lecture Notes in Computer Science. – 2001. – Vol.~2137. – P.~289–302.

ЭНТРОПИЙНЫЕ ХАРАКТЕРИСТИКИ ДВОИЧНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ В КРИПТОГРАФИИ

В. Ю. ПАЛУХА, Ю. С. ХАРИН

Современные средства криптографической защиты информации используют генераторы случайных и псевдослучайных последовательностей. Стойкость криптосистем зависит от того, насколько близка генерируемая последовательность по своим свойствам к равномерно распределённой случайной последовательности (РРСП). Одним из подходов к оценке качества генератора является статистическое оценивание энтропии и сравнение полученной оценки с ожидаемым значением для РРСП. Существуют различные функционалы энтропии, например, в [1] приводятся 23 функционала. Наибольшее распространение имеет энтропия Шеннона, которая и рассматривается в данной статье.

Статистическая оценка энтропии. Будем рассматривать стационарные в узком смысле двоичные последовательности $\{x_t\} \in V = \{0,1\}$ на некотором вероятностном пространстве (Ω, *F*, *P*). Пусть $p_{i_1,\dots,i_s} = P\{x_{t+1} = i_1, \dots, x_{t+s} = i_s\}$ – распределение вероятностей *s*-граммы $(x_{t+1},\dots,x_{t+s}) \in V_s$, которое не зависит от $t \in \mathrm{N}_0 = \mathrm{N} \cup \{0\}$. Многомерная (*s*-мерная) энтропия Шеннона для фрагмента длины $s \geq 1$ согласно [2] имеет вид

$$H\{x_1,\dots,x_s\} = h(s) = -\sum_{i_1,\dots,i_s \in V} p_{i_1,\dots,i_s} \ln p_{i_1,\dots,i_s}.$$

Пусть наблюдается n фрагментов длины s: $X^{(k)} = (x_1^{(k)},\dots,x_s^{(k)}) \in V_s, \quad k = 1,\dots,n$, сформированных из элементов последовательности $\{x_t\}$. Построим частотную оценку вероятности $p_J(s) = P\{X^{(k)} = J_1^s\}$, где $J_1^s = (j_1,\dots,j_s) \in V_s$ – мультииндекс, по n фрагментам длины $s \geq 1$:

$$\hat{p}_J(s) = \frac{v_J}{n}, \quad v_J = \sum_{k=1}^{n} \delta_{X^{(k)},J}, \quad \delta_{X^{(k)},J} = \begin{cases} 1, X^{(k)} = J; \\ 0, X^{(k)} \neq J \end{cases}$$

и затем построим оценку *s*-мерной энтропии по «подстановочному» принципу

$$\hat{h}(n,s) = -\sum_{J \in V_s} \hat{p}_J(s) \ln \hat{p}_J(s). \qquad (1)$$

Распределение вероятностей статистической оценки многомерной энтропии описывается следующей теоремой, которая опирается на результаты, полученные в [3].

Теорема. При истинной гипотезе H_* = {{x_t} есть РРСП} статистическая оценка *s*-мерной энтропии (1), построенная по подстановочному принципу, в специальной асимптотике $n, N = 2^s \to \infty, \quad n/N \to \lambda, \quad 0 < \lambda < \infty$, имеет асимптотически нормальное распределение: $\hat{h}(n,s) \sim N(\mu, \sigma^2)$, где для параметров асимптотического распределения справедливы следующие формулы