УДК 681.322

С.В. Беззатеев, Н.В. Волошина, К.А. Жиданов

Специальные классы кодов для стеганографических систем

Предлагается метод встраивания сообщения в контейнер, учитывающий значимость отдельных элементов исходного контейнера. Приведено сравнение предложенного метода и алгоритма F5 и его модификаций. Повышение эффективности встраивания достигается за счет использования специального класса кодов во взвешенной метрике Хэмминга, примененных к усовершенствованной модели стегоконтейнера.

Ключевые слова: стеганография, F5, коды во взвешенной метрике Хэмминга, WF5.

Стеганография – это скрытая передача информации по каналам связи – стегоканалам. При этом сообщения встраиваются в некий исходный объект (стегоконтейнер), который в дальнейшем открыто передается, в результате формируется скрытый канал передачи сообщения (стегоканал). Основной задачей в стегосистеме является обеспечение сокрытия факта передачи информации.

В современных системах инфокоммуникаций в качестве стегоканалов часто используют мультимедиаданные (изображения, звук), что обусловлено большой избыточностью данного вида информации. Для скрытой передачи часть избыточной информации заменяется на передаваемые сообщения.

Эффективность стеганографического метода определяется максимальным объемом встраиваемой информации, сложностью алгоритма встраивания и извлечения сообщений, а также устойчивостью к стегоанализу.

Обобщенная схема стегосистемы (этап встраивания) представлена на рис. 1. Извлечение информации производится в обратном порядке.

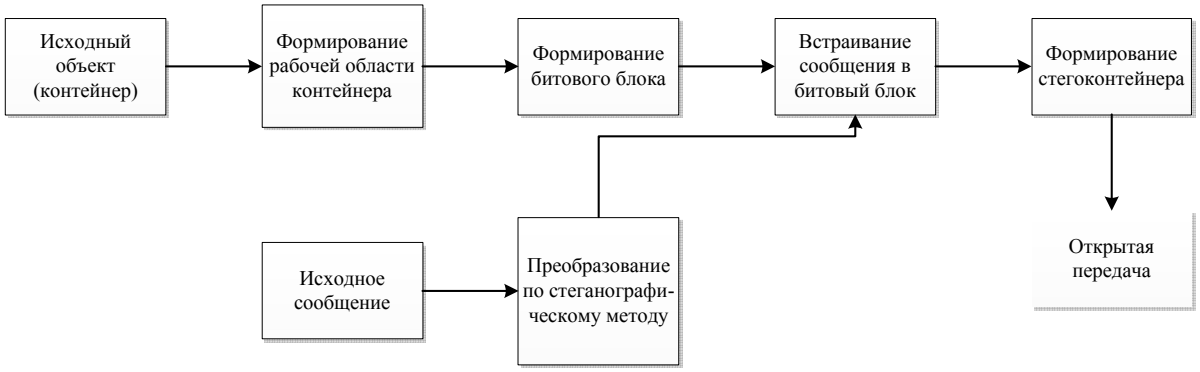

Рис. 1. Схема встраивания

Схема встраивания состоит из двух основных этапов:

- первый – формирование рабочей области контейнера и формирование битового блока;
- второй – процесс встраивания передаваемого сообщения в контейнер (формирование стегоконтейнера).

Задачей первого этапа является формирование рабочей области контейнера, куда будет производиться встраивание информации. Для современных методов характерен подход, при котором контейнер делится на две части (зоны), а при встраивании используется только одна из них. Например, встраивание с использованием одной зоны в мультимедиаданных, содержащей наименее значимую информацию. Наиболее известным примером являются LSB-методы встраивания в изображения [1, 2]. В результате рабочая область контейнера содержит биты, имеющие одинаковую значимость, с точки зрения влияния их изменения на искажение контейнера. Далее выделенная рабочая область преобразуется в блок бит (битовый блок), над которыми осуществляется преобразование встраивания.

Основной задачей на втором этапе является разработка метода встраивания передаваемого сообщения в сформированную рабочую область контейнера путем внесения изменений в сформиро-

ванный битовый блок. При этом решается задача обеспечения незаметности факта встраивания. Незаметность искажений принято оценивать отношением количества искаженных бит в битовом блоке к его объему. Чем большее значение имеет этот параметр, тем менее эффективным считается предлагаемый метод [3, 4]. Также эффективность метода встраивания определяется отношением объема встраиваемой информации к объему контейнера (рабочей области). Чем большее значение имеет данный параметр, тем эффективнее стегоалгоритм.

Такой подход упрощает модель исходного контейнера и первый этап схемы встраивания. Например, все множество элементов исходных мультимедиаданных рассматривается как множество элементов двух типов: значащие и незначащие. Это очень ограничивает объем рабочей области контейнера, а следовательно, и эффективность стегосистемы.

В статье предлагается новая модель формирования стегоконтейнера (в качестве примера, здесь рассматривается случай, когда мультимедиаданные представляют собой изображение), предполагающая наличие в исходных мультимедиаданных нескольких множеств элементов (зон), имеющих различную значимость (вес). Значимость определяется влиянием искажений элементов зон на заметность вносимых искажений для исходных мультимедиаданных. Такая модель позволяет расширить объем рабочей области контейнера, формируемой на первом этапе. При этом на втором этапе предлагается использовать метод встраивания сообщения в контейнер, учитывающий веса выделенных на первом этапе зон.

Одним из эффективных методов встраивания информации в контейнер является использование помехоустойчивых кодов [3]. В статье рассматривается возможность повышения эффективности стегоалгоритма с учетом предложенной взвешенной модели контейнера и согласованных с ней специальных классов помехоустойчивых кодов во взвешенной метрике Хэмминга [5].

Стеганографический метод F5

В качестве базового метода встраивания сообщений, использующего помехоустойчивые коды, рассмотрим алгоритм F5 [3], структурная схема которого представлена на рис. 2.

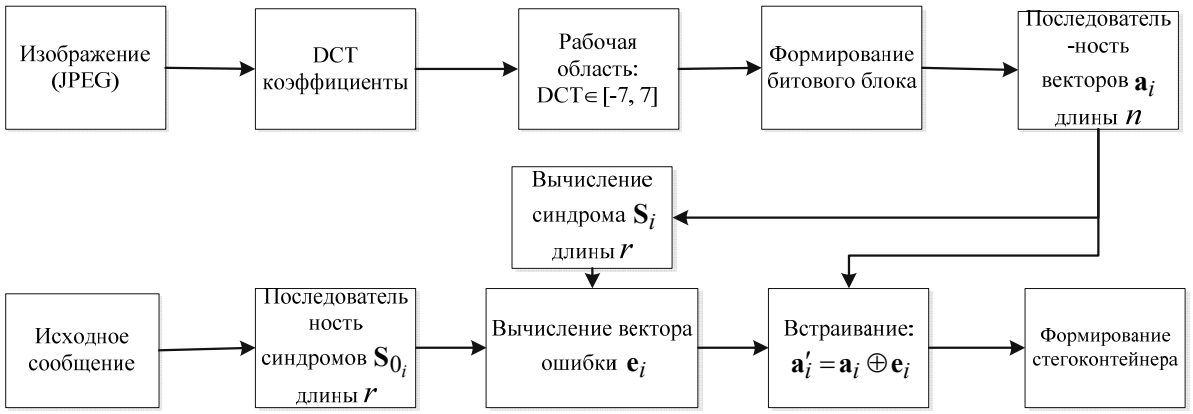

Рис. 2. Схема F5

В качестве исходных мультимедиаданных будем рассматривать изображение в формате JPEG. Для формирования рабочей области контейнера в F5 предлагается использовать коэффициенты DCT преобразования малых величин. Для встраивания применяют кодовую конструкцию, использующую код Хэмминга.

Как известно, код Хэмминга является совершенным, что обеспечивает соответствие любому ненулевому двоичному вектору длины r (синдрому) уникального вектора $\mathbf{e}$ (вектор ошибки) длины n с числом единиц, равным 1, т.е. вес Хэмминга такого вектора $wt_H(\mathbf{e})=1$. Проверочная матрица $\mathbf{H}$ такого кода имеет в качестве столбцов все возможные ненулевые векторы длины r. Очевидно, что число таких векторов $n=2^r-1$. Например, для $r=3$ матрица $\mathbf{H}$ имеет следующий вид:

$$\mathbf{H}=\begin{bmatrix}0&0&1&1&0&1&1\\0&1&0&1&1&0&1\\1&0&0&0&1&1&1\end{bmatrix}.$$

Синдром $\mathbf{S}$, вектор $\mathbf{a}$ и проверочная матрица кода Хэмминга $\mathbf{H}$ связаны соотношением

$$\mathbf{a}\cdot\mathbf{H}^T=\mathbf{S}\,.$$

Описание процедуры встраивания стеганографического метода F5:

- Встраиваемая информация I разбивается на последовательности длины r бит.
- Данные последовательности рассматриваются как синдромы $\mathbf{S}_{0_i}$, полученные по заданной проверочной матрице $\mathbf{H}$ и соответствующим вектором $\mathbf{e}_i$ длины n.
- Рабочая область контейнера (битовый блок) рассматривается как последовательность векторов $\mathbf{a}_i$ длины n с синдромами $\mathbf{S}_i$, полученными по той же проверочной матрице $\mathbf{H}$.
- Процедура встраивания состоит в замене текущего вектора $\mathbf{a}_i$ длины n битового блока на вектор $\mathbf{a}'_i=\mathbf{a}_i\oplus\mathbf{e}_i$, дающий требуемый синдром $\mathbf{S}_{0_i}$, равный передаваемой последовательности длины r бит.
- Для F5 такая замена возможна путем внесения искажений $\mathbf{e}_i$, $wt_{\mathrm{H}}\left(\mathbf{e}_i\right)\le 1$ в битовый блок рабочей области контейнера.

Работа F5 схематически представлена на рис. 3.

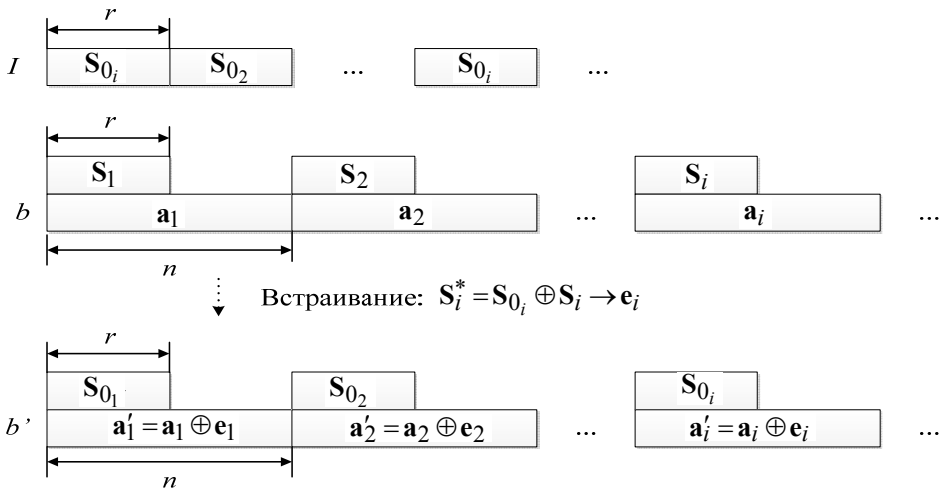

Рис. 3. Схема процедуры встраивания в F5

Для данного метода в дальнейшем были предложены идеи усовершенствования за счет использования других кодов, например БЧХ-кодов, исправляющих две ошибки [6]. В результате при увеличении максимального числа вносимых искажений увеличилось соотношение длины синдрома r к длине вектора n, что значительно увеличило максимальный объем встраиваемой информации. В то же время отношение числа искаженных бит к объему рабочей области стало хуже, поскольку для выбранных кодов БЧХ максимальное число искажений t на длину вектора n равно 3.

В представленных выше методах для формирования рабочей области контейнера используется примитивная модель контейнера, содержащая только две зоны. В F5 в качестве рабочей области используются DCT коэффициенты, значения которых принадлежат множеству $\{-7,-6,\ldots,6,7\}\setminus\{0\}$. Внесение искажений в F5 реализуется с помощью изменения знака коэффициента, соответствующего изменяемому биту в битовом блоке. Рассматриваемые коэффициенты относят к зоне наименее значимых значений, влияние искажений в которых считается незначительным и незаметным. Такое ограничение при формировании рабочей области контейнера уменьшает эффективность стегоалгоритма вследствие значительного сужения рабочей области (длины битового блока) контейнера. Рабочую область можно расширить за счет использования зон с более высокой значимостью, что позволяет повысить эффективность работы алгоритма встраивания [7].

Предлагаемый метод

При формировании рабочей области контейнера в рассмотренных алгоритмах исходное изображение (контейнер) разделено всего на две зоны: значимую зона и незначимую. Для значимой зоны не допускаются никакие искажения, а для незначимой искажения допустимы в любой ее части. Именно незначимая зона исходного изображения и выбирается в качестве рабочей области. Так, в

результате работы алгоритма F5 вносимые искажения равномерно распределены между коэффициентами рабочей зоны.

Взвешенная модель контейнера

Известно, что в изображении возможно выделение нескольких зон значимости. Этот факт широко используется в алгоритмах сжатия изображений с потерями. Например, при сжатии JPEG используется тот факт, что чем больше значение DCT-коэффициента, тем выше его значимость, а также учитывается позиция DCT-коэффициента в блоке матрицы DCT-преобразования. Следовательно, контейнер может иметь более сложную взвешенную структуру. Для более значимых элементов допускается меньше искажений, чем для менее значимых.

Таким образом, элементам рабочей области контейнера назначаются веса w_j, задающие их значимость. Вес элемента w_j характеризует вклад искажений элементов контейнера данного веса w_j в заметность результирующих искажений контейнера. Элементы одинакового веса объединяются в зоны. В результате контейнер представляет собой множество зон $\{z_1, z_2, \ldots, z_m\}$, для каждой из которых определен соответствующий вес значимости $\left\{w^{z_1}, w^{z_2}, \ldots, w^{z_m}\right\}$ (рис. 4).

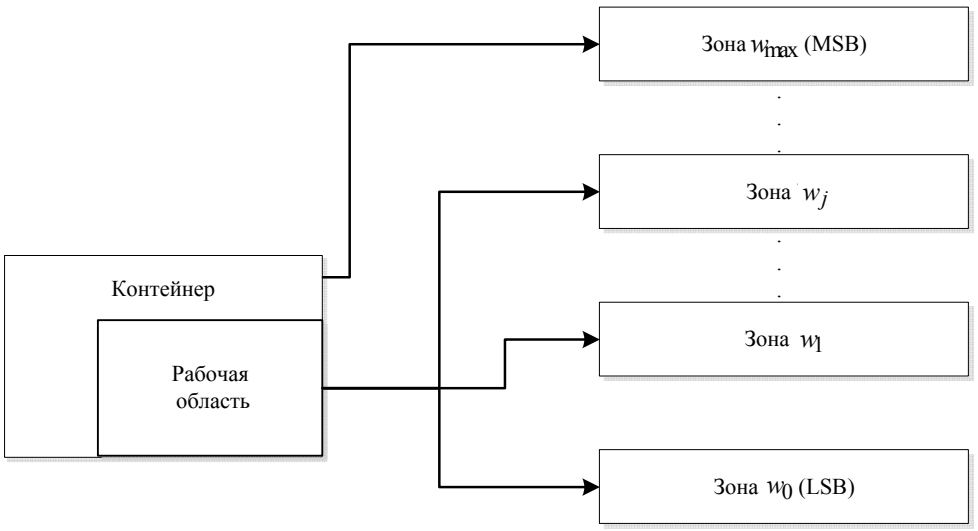

Рис. 4. Формирование взвешенной рабочей области

Значения весов $w^{z_1}, w^{z_2}, \ldots, w^{z_m}$ определяются типом используемого контейнера, а также типом его элементов. В статье, в качестве примера, рассматривается тип контейнера – полутоновое изображение, а тип элементов – коэффициенты DCT преобразования.

В случае представления контейнера, как множества элементов, разбитого на m зон с различной значимостью (весом w^{z_j}), можно сформировать рабочую область контейнера с взвешенными элементами. Для встраивания информации в такой взвешенный контейнер предлагается использовать коды во взвешенной метрике Хэмминга [5]. Такой подход позволяет согласовать метод встраивания с взвешенной моделью контейнера.

Коды во взвешенной метрике Хэмминга

В большинстве случаев в теории кодирования рассматривается ситуация, когда каждая позиция передаваемого вектора **a** длины *n* имеет одинаковую вероятность p искажения. То есть, в двоичном случае для каждой позиции передаваемого вектора **a** вероятность перехода из 1 в 0 и из 0 в 1 равна p. Однако существуют модели каналов передачи и хранения информации, в которых вероятность искажения на различных позициях различна. Тогда длина вектора представляется суммой длин $n_1, n_2, \ldots, n_l : n = \sum_{i=1}^{l} n_i$. Вероятность искажения символа, соответственно, различна – $p_1, p_2, \ldots, p_l$ и зависит от того, на какой позиции он находится. Очевидно, что в этом случае эффективным решением является использование помехоустойчивых кодов во взвешенной метрике Хэмминга. Такие коды позволяют адаптировать корректирующую способность кода к неравномерному

распределению вероятности ошибки на длине передаваемого вектора . Задание таких кодов определяется вектором весов позиций $W=\{w_1,w_2,\ldots,w_l\}$, набором длин $\{n_1,n_2,\ldots,n_l\}$, числом информационных символов k и корректирующей способностью кода, определяемой его расстоянием во взвешенной метрике Хэмминга:

$$\frac{d-1}{2}\geq w_1\cdot\tau_1+w_2\cdot\tau_2+\ldots+w_l\cdot\tau_l,$$

где $t=\tau_1+\tau_2+\ldots+\tau_l$ – общее число ошибок, которые такой код может исправить, τ_i – число ошибок, исправляемых кодом на n_i позициях, имеющих вес w_i и вероятность ошибки p_i. Максимальное значение τ_i определяется как

$$\tau_{i_{\max}}=\left\lfloor\frac{d-1}{w_i\cdot 2}\right\rfloor.$$

Описанные выше коды хорошо согласуются с предложенной взвешенной моделью контейнера и могут быть использованы в новом стеганографическом алгоритме, построенном на базе F5.

В стандартном методе F5 используется самая простая модель контейнера, состоящая всего из двух зон значимости: z_0 – LSB и z_1 – MSB с весами w^{z_0} и w^{z_1} соответственно. Рабочая область контейнера состоит только из элементов зоны z_0. Таким образом, встраивание производится только за счет изменения DCT-коэффициентов, значения которых принадлежат множеству $\{-7,-6,\ldots,6,7\}/\{0\}$. Для расширения рабочей области контейнера воспользуемся описанной выше взвешенной моделью контейнера. Допустим, контейнер описывается моделью, в которой существуют три взвешенные зоны значимости z_0^w, z_1^w и z_2^w с весами соответственно w^{z_0},w^{z_1},w^{z_2}. Пусть зона z_0^w полностью совпадает с зоной z_0. А зона z_1 разделена на две: z_1^w и z_2^w. Зона z_2^w является MSB-зоной и не допускает внесения искажений. Зона z_1^w выделена из зоны z_1 и содержит элементы контейнера, допускающие искажения, но в меньшей степени, чем элементы зоны z_0. Тогда для элементов зоны z_1^w назначаются веса большие, чем веса элементов зоны z_0^w, т.е. $w^{z_0}<w^{z_1}<w^{z_2}=\infty$. Такое представление позволяет расширить размер рабочей зоны контейнера.

При использовании взвешенной модели в процессе формирования битового блока также необходимо учитывать веса элементов, преобразуемых в биты. В этом случае битовый блок представляет собой набор взвешенных подблоков b_i с разными весами w_i^b. Искажения бит разных подблоков оказывают неодинаковое влияние на заметность искажений в результирующем стегоконтейнере. Для алгоритма F5 и его модификаций неравнозначность весов подблоков можно задавать максимальным числом искажений t на длине вектора $\mathbf{a}$. Для кода Хэмминга эта величина равна 1 [3], а для БЧХ равна 3 [6]. В работе [6] показано, что для стандартной рабочей области F5 z_0 допускается значение t=1. Тогда и для зоны z_0^w допустимо значение $t=3$. Для более значимой зоны z_1^w значение для t равно 1. Тогда для внедрения информации в первом подблоке b_0 выбирается БЧХ (BCH), код, исправляющий две ошибки, а для b_1 – код Хэмминга. При этом встраивание информации в подблоки происходит независимо.

Предлагаемый нами вариант, учитывающий веса, назначенные блокам, в качестве инструмента встраивания, использует коды во взвешенной метрике Хэмминга.

Тогда в случае невзвешенного встраивания для векторов $\mathbf{a}_{b_0}$ битового подблока b_0 используются следующие параметры: $n_{b_0}=15,\ t_{b_0}=3,\ r_{b_0}=8$. Для векторов $\mathbf{a}_{b_1}$ подблока b_1 используются: $n_{b_1}=7,\ t_{b_1}=1,\ r_{b_1}=3$. В этом случае для итогового вектора $\mathbf{a}_b$ битового блока используются параметры: $n_b=22,\ t_b=4,\ r_b=11$. Такой способ встраивания обозначим BCH+F5.

В случае взвешенного встраивания с такими же параметрами взвешенного битового блока b^w для вектора $\mathbf{a}_b^w$ задаются параметры: $n_b^w=n_{b_0}^w+n_{b_1}^w=15+7=22$, $t_b^w=4$, $r_b^w=8$. При данных параметрах для подблоков взвешенного вектора $\mathbf{a}_b^w$ назначаются веса $w_{b_0}=1$ и $w_{b_1}=4$. Такой способ встраивания обозначим как WF5.

Для описанного выше варианта распределения зон и весов была построена схема встраивания, представленная на рис. 5.

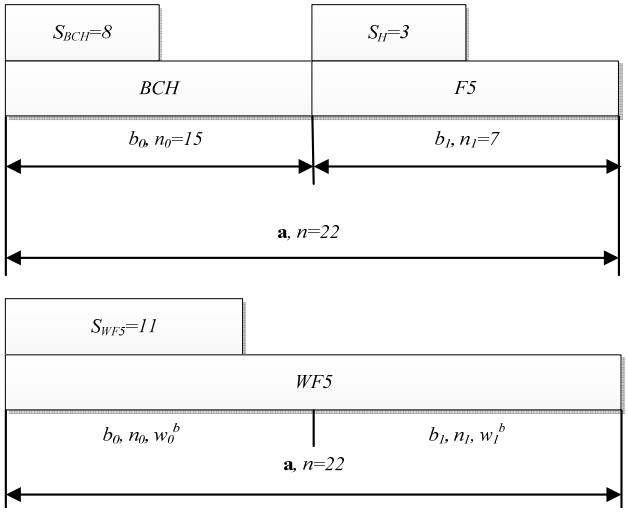

Рис. 5. Пример стеганографического встраивания с использованием взвешенной модели контейнера

Сравнение эффективности методов производится по параметрам n – длина вектора встраивания **a**, t – максимальное число искажений на длине n, r – длина синдрома вектора **a**, $\hat{t}$ – оценка среднего числа искажений на длине n.

Для сравнения рассматриваемых методов с учетом введенной взвешенной модели контейнера введем дополнительный параметр – значение штрафную функцию. Функция F учитывает неравнозначность влияния искажений, вносимых в разных зонах контейнера, на их заметность в стегоконтейнере. Функция F добавляет штраф за появление искажений в зависимости от веса зоны. Чем больше вес зоны, тем больше вес вносимого искажения. Таким образом, штрафная функция имеет следующий вид:

$$F(\eta) = \sum_{i=0}^{l} t_i \cdot \eta_i,$$

где η_i – штрафной коэффициент искажения, вносимого в i-й зоне контейнера.

Результаты сравнения эффективности методов представлены в таблице.

Сравнение алгоритмов

Параметр оценки эффективности	Алгоритм	
	F5+BCH	WF5
Длина вектора **a**	22	22
Длина синдрома r	8+3	11
Максимальное число искажений t на длине вектора **a**	3+1	4
Оценка средней ошибки $\hat{t}$ на длине вектора **a**	3,34	3,44
$F(\eta_0 = 1,\ \eta_1 = 4)$	5,97	3,63
$F(\eta_0 = 1,\ \eta_1 = 2)$	4,22	3,62

График зависимости изменения значения штрафной функции $F(\eta)$ от величины штрафа η_1, характеризующего попадание вносимого искажения в зону z_1^w, для взвешенного метода WF5 (жирная линия) и невзвешенного метода BCH+F5 (пунктирная линия) представлен на рис. 6.

Из представленного графика видно, что если не учитывать значимости искажений $(\eta_0 = \eta_1 = 1)$, то метод BCH+F5 является более эффективным. Однако при увеличении значения штрафа η_1 за попадание искажений в более значимую зону z_1 эффективность предложенного взвешенного метода WF5 возрастает. Таким образом, для взвешенных моделей контейнера предложенный метод дает большую эффективность, чем стандартные методы стеганографического встраивания.

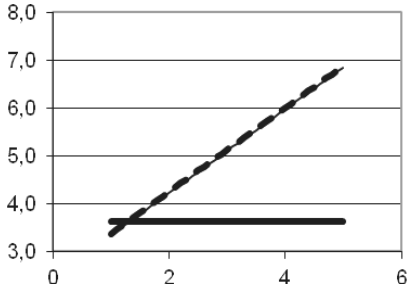

Рис. 6. Оценка эффективности встраивания по штрафной функции $F(\eta)$

Заключение

Использование предложенной в статье новой модели контейнера, согласованной со свойствами конкретного типа контейнера, позволило разработать новый стеганографический алгоритм, учитывающий значимость отдельных зон исходного контейнера и использующий коды во взвешенной метрике Хэмминга, согласованные с выбранным набором весов. Представленные в статье результаты показывают эффективность предложенного метода по сравнению с алгоритмом F5 и его извест-

ными модификациями. Открытой остается задача поиска оптимального соотношения параметров для предложенного метода, а именно:

1. Поиск весовых функций для различных типов контейнеров.
2. Оптимизация процедуры построения рабочей зоны взвешенного контейнера.
3. Построение согласованных с взвешенным контейнером кодов во взвешенной метрике Хэмминга.
4. Оптимизация параметров и вида штрафной функции для оценки искажений при использовании предложенного подхода.
5. Расширение предложенного подхода с использованием взвешенной модели контейнера на другие классы стеганографических систем.

Литература

1. Chandramouli R. Analysis of LSB based image steganography techniques // Image Processing: International Conference. – 2001. – P. 1019–1022.

2. Neeta D. Implementation of lsb steganography and its evaluation for various bits / D. Neeta, K. Snehal, D. Jacobs // Digital Information Management: 1st International Conference – 2006.– P. 173–178.

3. Westfeld A. High Capacity Despite Better Steganalysis (F5-A Steganographic Algorithm) / I.S. Moskowitz (eds.) // Information Hiding. 4-th International Workshop. Lecture Notes Computer Science. – Berlin; Heidelberg; New York: Springer-Verlag, 2001. – Vol. 2137. – – P.. 289–302.

4. Galand F. Information hiding by coverings / F. Galand, G. Kabatiansky // Proc. of IEEE Information Theory Workshop. – 2003. – P. 151–154.

5. Bezzateev S. Class of binary generalized Goppa codes perfect in weighted Hamming metric / S. Bezzateev, N. Shekhunova // Worrkshop on coding and cryptography. – Paris, France, 2011. – P. 233–242.

6. Medeni M.B. A Novel Steganographic Protocol from Error-correcting Codes / M.B. Medeni, El.M. Soudi // Journal of Information Hiding and Multimedia Signal Processing. – 2010. – P. 339–343.

7. Bezzateev S. Special Class of (L,G) Codes for Watermark Protection in DRM / S. Bezzateev, N. Voloshina, V. Minchenkov // Eighth International Conference on Computer Science and Information Technologies. – Yerevan, Armenia, 2011. – P. 225–228.

Беззатеев Сергей Валентинович
Д-р техн. наук, доцент, зав. каф. технологий защиты информации
Санкт-Петербургский государственный университет аэрокосмического приборостроения (ГУАП)
Тел.: 8 (812) 494-70-52
Эл. почта: bsv@aanet.ru

Волошина Наталия Викторовна
Канд. техн. наук, доцент каф. технологий защиты информации ГУАП
Тел.: 8 (812) 494-70-52
Эл. почта: natali@vu.spb.ru

Жиданов Константин Александрович
Ассистент каф. технологий защиты информации ГУАП
Тел.: 8 (812) 494-70-52
Эл. почта: konstantin.zhidanov@gmail.com

Bezzateev S.V., Voloshina N.V., Zhidanov K.A.

Special class of error correcting codes for steganographic systems

A new method for embedding messages in the container, taking into account the significance of the original container elements, is proposed. The comparison of the proposed method and algorithm F5 and its modifications is represented. Enhancement of the stegoalgorithm efficiency is achieved by using a special class of codes in the weighted Hamming metric applied to the improved weighed model of stegocontainer.

Keywords: steganography, F5, Hamming waited metric codes, WF5.