УДК [004.056+003.26](075.8)

КОРРЕКЦИЯ ОДИНОЧНЫХ И ДВОЙНЫХ ПАРНЫХ ОШИБОК В СТЕГАНОГРАФИЧЕСКИХ КАНАЛАХ ПЕРЕДАЧИ ИНФОРМАЦИИ

П.П. Урбанович[1, 2]

[1]*Белорусский государственный технологический университет,*
ул. Свердлова, 13а, 220005 Минск, Беларусь, p.urbanovich@belstu.by
[2]*Люблинский Католический университет им. Яна Павла II,*
al. Racławickie 14, 20-950 Lublin, Poland, pavel.urbanovich@kul.pl

Проанализированы особенности использования избыточных корректирующих кодов в стеганографических приложениях. Приведено формальное описание модели стеганосистемы, в которой код используется для коррекции извлекаемого из стеганоконтейнера сообщения. Предложена конструкция линейного кода для коррекции одиночных и двойных парных (смежных) ошибок.

Ключевые слова. Линейный корректирующий код; парная ошибка; стеганография; скрытый канал.

CORRECTION OF SINGLE AND DOUBLE ERRORS IN STEGANOGRAPHIC CHANNELS OF INFORMATION TRANSMISSION

P.P. Urbanovich[1, 2]

[1]*Belarusian State Technological University,*
Sverdlova str., 220005 Minsk, Belarus, p.urbanovich@belstu.by
[2]*The John Paul II Catholic University of Lublin,*
Al. Racławickie 14, 20-950 Lublin, Poland, pavel.urbanovich@kul.pl

The features of the use of redundant correcting codes in steganographic applications are analyzed. A formal description of the steganosystem model, in which the code is used to correct the message extracted from the steganocontainer, is given. A construction of a linear code for correction of single and double paired (adjacent) errors is proposed.

Keywords. Linear error-correcting code; pair error; steganography; covert channel.

Введение

Стеганографические алгоритмы позволяют скрывать конфиденциальную информацию в каналах, формируемых носителями другой информации, или стеганографическими контейнерами (c). Такие каналы называют скрытыми. Исходная скрываемая информация (m), помимо основных стеганографических трансформаций (осаждение/извлечение), может

подвергаться также другим канальным преобразованиям: например, на основе криптографии и/или помехоустойчивого кодирования. Перечисленные преобразования информации на обеих сторонах канала можно соотнести с многоключевой стеганосистемой, основу которой составляет скрытый канал или стеганосообщение (s): сообщение + контейнер [1].

Идея использования помехоустойчивых кодов совместно со стеганографией в наиболее общем виде изложена в [2]; здесь автор назвал метод кодирования матричным. Одна из практических реализаций матричного кодирования связана с известным стеганографическим алгоритмом F5 [3]. Дальнейшие исследования в указанной предметной области преследовали по существу одну основную цель: обеспечить минимальные искажения c при осаждении исходного сообщения m, ориентируясь, как правило, на алгоритмы из класса LSB (List Sygnificant Bits – наименее значащих битов). Для решения этой задачи были созданы коды для записи на «мокрой бумаге», называемые также кодами «мокрой бумаги» (Wet Paper Codes, WPC) [4], комбинации кодов Хемминга и WPC [5], синдромные решетчатые коды (Syndrome Trellis Codes, STC) [6], стеганографические полярные коды (Steganographic Polar Codes, SPC) [7] и др.

Адаптация битов сообщения m из множества M под содержание соответствующих битовых последовательностей контейнера c из множества C на основе перечисленных или подобных им корректирующих кодов повышает стеганостойкость системы передачи или хранения скрытого сообщения. В данном случае под стеганостойкостью понимается степень модификации исходного содержания c после размещения в нем m, т. е. после преобразования контейнера c в стеганоконтейнер s: $c \rightarrow s$ ($s \in S$, S – множество всех возможных типов контейнеров).

Однако не менее важной является проблема обеспечения целостности осажденного в контейнер сообщения m после случайной или преднамеренной модификации стеганоконтейнера s. Нами, в частности, установлено, что простейшие конвертации s (docx–pdf–docx) в случаях использования c в виде текстового документа при реализации метода LSB на основе модификации битов цветовых каналов модели RGB при осаждении битов сообщения m в некоторых случаях приводит в конечном итоге к изменению цветового кода (в границах от 0 до 255). Если при осаждении m модифицируются два младших бита (из восьми) во всех или в отдельных цветовых кодах (R, G, B) пикселей, составляющих c, то с наибольшей вероятностью в результате указанных конвертаций значения одного или обоих битов, составляющих часть m, могут измениться. Это означает, что формальное воздействие на скрытый канал помехи (в виде указанной конвертации или в ином проявлении) приводит к появлению одиночной, а также двойной смежной (или парной)

ошибки. Практически единственным противодействием ошибкам является корректирующий код.

Специфика ошибок диктует необходимость поиска такой кодовой конструкции, которая обеспечивала бы положительный эффект в сравнении с применением известных кодов, обнаруживающих и корректирующих одиночные и двойные независимые ошибки в кодовых словах подобно тому, как может решаться задача повышения функциональной надежности систем и устройств полупроводниковой памяти [8]. Рассмотрению и анализу такого кода и посвящена настоящая работа.

1. Основная часть.

1.1. Формальное описание стеганографической системы и ее элементов.

Рассматриваемая стеганографическая система S с использованием корректирующего кода в наиболее общем виде может быть представлена следующим образом [1]:

$$S = (M, C, S, K, K_д, F, F^{-1}), \quad (1)$$

где $M, C, S, K, K_д$ – соответственно конечные множества, содержащие: возможные тайные сообщения M ($m \in M$); используемые контейнеры ($c \in C$); стеганоконтейнеры ($s \in S$); основные ключи (K, $k \in K$; k – отдельно взятый ключ), относящиеся к используемым стеганографическим методам (например, LSB) осаждения/извлечения тайного сообщения; дополнительные ключи ($K_д$, $k_д \in K_д$; $k_д$ – отдельно взятый дополнительный ключ), относящиеся к модификациям основного стеганометода (например, количество младших битов кодов R, G, B, используемых для внедрения m; или порядок выбора элементов c – последовательно или по иному принципу – для размещения битов m). При этом полагаем, что конкретное сообщение m в двоичном виде состоит из t битов: $m = m_1, m_2, \ldots, m_t$, которые разделяются на b блоков по l битов в каждом: $t = bl$.

Последние элементы в правой части (1): F, F^{-1} – функциональные преобразования (отображения), соответственно обозначающие внедрение закодированных на основе избыточного кода сообщений в контейнер и обратные функциональные преобразования: извлечение закодированных сообщений и исправление в них обнаруженных ошибок. В общем случае

$$F: M \times f_{ECC} \times C \times K \times K_д \rightarrow S, \quad (2)$$

$$F^{-1}: S' \times K \times K_д \rightarrow M' \times f_{ECC}, C. \quad (3)$$

В последних выражениях произведение $M \times f_{ECC}$ означает выполнение операции избыточного кодирования каждого из блоков сообщений M на

основе кода, корректирующего ошибки (Error Correcting Code, ECC); произведение $S' \times K \times K_д$ – функцию извлечения сообщений M' из стеганоконтейнеров S', $M' \times f_{ECC}$ – декодирование извлеченных сообщений M' с обнаружением и исправлением ошибок. Полагаем, что различные вышеупомянутые воздействия на скрытый канал стеганосистемы **S** (стеганоконтейнер S, $s \in S$) проводят к модификации S: $S \rightarrow S'$, $S \neq S'$, соответственно $s \neq s'$ ($s' \in S'$) и $M \neq M'$; в конечном итоге $m \neq m'$, где $m' = (m')_1$, $(m')_2, \ldots, (m')_t$.

1.2. Избыточный помехоустойчивый код, корректирующий одиночные и двойные парные ошибки.

Дальнейший анализ будем производить, опираясь на общепринятые положения теории избыточного кодирования информации (см., например, [9]).

Любой корректирующий код задается тремя основными параметрами: длиной кодируемого или информационного слова (в нашем случае – l), общей длиной кодового слова или блока (обозначим его символом n) и минимальным расстоянием Хемминга (d) между двумя кодовыми словами: (n, l, d)-код. Отдельный блок исходного сообщения m обозначим $\dot{m}_L$ (m состоит из b таких блоков); $\dot{m}_L = \dot{m}_1, \dot{m}_2, \ldots, \dot{m}_l$ и $\dot{m}_i = \{1, 2\}$, $i = 1, \ldots, l$. Закодированное сообщение $\dot{m}_L$ состоит из n символов. Это закодированное сообщение обозначим $\dot{m}_N$ ($\dot{m}_N = \dot{m}_1, \dot{m}_2, \ldots, \dot{m}_l, \dot{m}_{l+1}, \dot{m}_{l+2}, \dot{m}_{l+r}$; $n=l+r$).

Избыточность кода – $R = n/l$. Применительно к одному блоку внедряемого в контейнер s сообщения кодовое слово $\dot{m}_L$ является результатом выполнения операции $M \times f_{ECC}$ в выражении (2).

Подобным образом будем обозначать части извлекаемого из стеганоконтейнера s' закодированного блока (обозначим его $\dot{m}_N'$; $\dot{m}_N' = \dot{m}_1', \dot{m}_2', \ldots, \dot{m}_l', \dot{m}_{l+1}', \dot{m}_{l+2}', \dot{m}_{l+r}'$) сообщения.

Основная часть нашего исследования и анализа состоит в разработке такой конструкции линейного блочного (длина блока или кодируемого сообщения $\dot{m}_L$ равна l битам) корректирующего кода, который позволял бы обнаруживать и корректировать одиночные и двойные ошибки в смежных символах кодового слова $\dot{m}_N'$. Эта операция использования кода формально описывается частью выражения (3): $M' \times f_{ECC}$.

Линейный код Хемминга однозначно задается с помощью проверочной матрицы Хемминга, $\boldsymbol{H}_{r \times n}$ размерностью $r \times n$:

$$\boldsymbol{H}_{r \times n} = [\, \mathbf{I} \mid \mathbf{A} \,], \qquad (4)$$

где **I** – диагональная матрица размерностью $r \times r$, **A** – матрица размерностью $r \times l$, вес Хемминга каждого вектор-столбца $\mathbf{a}_j$ ($j = 1, 2, \ldots, l$) не менее двух: $wt(\mathbf{a}_j) \geq 2$.

При фиксированных l корректирующие свойства кода определяются параметром $R = (l+r)/l$, т.е. количеством r избыточных символов, которые нужно вычислить и присоединить к информационному слову $\dot{m}_L$.

Наши рассуждения при конструировании кода с заданными корректирующими способностями будем строить на следующих простых положениях и оценках:

а) избыточные символы вычисляются на основе соотношения

$$\boldsymbol{H}_{r\times n} \times (\dot{m}_N)^{\mathrm{T}} = 0, \tag{5}$$

здесь «т» означает транспонирование;

б) в системе применяется синдромный метод декодирования извлеченного из s' кодового слова $\dot{m}_N'$;

в) синдром ошибки Sr – r-разрядный вектор, вес Хемминга которого равен нулю при отсутствии ошибок в $\dot{m}_N'$ ($\dot{m}_N' = \dot{m}_N$); при этом в общем случае

$$Sr = \boldsymbol{H}_{r\times n} \times (\dot{m}_N')^{\mathrm{T}}, \tag{6}$$

и в соответствии с (5), (6) Sr равен сумме (mod 2) тех вектор-столбцов $\boldsymbol{H}_{r\times n}$, позиции которых соответствуют местоположению ошибок в $\dot{m}_N'$;

г) в извлеченном ($M'\times f_{\mathrm{ECC}}$) кодовом слове $\dot{m}_N'$ длиной n битов могут появиться n одиночных ошибок (в любом отдельно взятом символе) и $\lceil n/2 \rceil$ двойных парных (смежных) ошибок: в битах 1-2, 3-4, 5-6 и т. д.;

д) последнее положение – с учетом п. б) – означает, что конструкция матрицы **A** должна быть такой, чтобы суммы по модулю два смежных парных вектор-стобцов были разными и имели вес Хемминга больший 1(например, $wt(\mathbf{a}_1+\mathbf{a}_2) \geq 1 \bmod 2$ или $wt(\mathbf{a}_3+\mathbf{a}_4) \geq 1 \bmod 2$ и т.д.);

е) с учетом сформулированных положений должно выполняться следующее условие:

$$2^r \geq 1 + n + n/2$$

или иначе

$$2^r \geq 1 + (l+r) + (l+r)/2. \tag{7}$$

Решение неравенства (5) относительно r дает такой результат:

$$r \geq \log_2 l + 2. \tag{8}$$

Последнее выражение характеризует разрабатываемую конструкцию кода как приближение по параметру относительной избыточности (R) к коду Хемминга с минимальным кодовым расстоянием d=4.

2. Результаты и их обсуждение.

В статье представлен новый механизм использования избыточного кода совместно со стеганографическим преобразованием информации на основе, например, метода LSB. Основная особенность этого механизма заключается не в адаптации закодированной тайной информации к содержанию контейнера, а в защите ее целостности. Предложенная конструкции кода для обнаружения и коррекции одиночных и двойных парных (в смежных символах) ошибок характеризуется меньшей избыточностью по сравнению с кодами для коррекции одиночных и двойных независимых ошибок. Классическим примером последних являются коды БЧХ [9, с.87–95, с.566]. Для сравнения в таблице приведены соответствующие характеристики двух кодов. В таблице используется следующий формат записи данных: БЧХ | ПК. Видно, что предложенная конструкция кода по количеству избыточных символов r для фиксированного l обладает преимуществом перед БЧХ.

Таблица – Сравнение параметров предложенного кода (ПК) и кода БЧХ

l	4	8	16
r	6 | 4	8 | 5	10 | 6
n	10 |8	16 | 13	26 | 22
R	2,5 | 2,0	2,0 | 1,625	1,625 | 1,373

Ниже также приведены примеры проверочных матриц предложенных кодов (8, 4): для $n = 4$ $r = 4$ и (13, 5) – $n = 13$ $r = 5$. В первом случае информационное слово состоит из полубайта, во втором – из 8 битов, что соответствует кодированию одного символа текста в кодах ASCII:

$$\boldsymbol{H}_{4\times 8} = \begin{matrix} 1101\ 1 \\ 0111\ \ 1 \\ 1111\ \ \ 1 \\ 0110\ \ \ \ 1, \end{matrix} \qquad \boldsymbol{H}_{5\times 13} = \begin{matrix} 11011100\ 1 \\ 11110110\ \ 1 \\ 01101111\ \ \ 1 \\ 11110001\ \ \ \ 1 \\ 01001011\ \ \ \ \ 1\ . \end{matrix}$$

Отмечаем, что если пары столбцов считать слева направо, то одиночным и парным ошибкам соответствуют не повторяющиеся синдромы.

Библиографические ссылки

1. Urbanovich P., Shutko N. Theoretical Model of a Multi-Key Steganography System // Recent Developments in Mathematics and Informatics. Contemporary Mathematics and Computer Science. Vol. 2, Chapter 11. Lublin: KUL, 2016. P. 181–202.
2. Crandall R. Some notes on steganography. Posted on steganography mailing list. 1998. URL: http://os.inf.tu-dresden.de/~westfeld/crandall.pdf .
3. Westfeld A. F5: a steganographic algorithm // Proc. 4th Int. Workshop Information Hiding 2001, Lecture Notes in Computer Science, vol. 2137. 2001. P. 289–302.

4. Fridrich J., Goljan M. and Soukal D. Efficient wet paper codes. In Proceedings of Information Hiding, Springer Verlag, 2005.
5. Zhang, W., Zhang, X., Wang, S. Maximizing Steganographic Embedding Efficiency by Combining Hamming Codes and Wet Paper Codes. In: Solanki, K., Sullivan, K., Madhow, U. (eds) Information Hiding. IH 2008. Lecture Notes in Computer Science, vol 5284. Springer, Berlin, Heidelberg, 2008. https://doi.org/10.1007/978-3-540-88961-8_5.
6. Filler T., Judas J., Fridrich J. Minimizing additive distortion in steganography using syndrome-trellis codes // IEEE Transactions on Information Forensics and Security. 2011 № 6(3-2). P. 920–935.
7. Li W., Zhang W., Li L., Zhou H., Yu N. Designing near-optimal steganographic codes in practice based on polar codes // IEEE Transactions on Communications. 2020. № 68(7). P. 3948–3962.
8. Урбанович П.П., Алексеев В.Ф., Верниковский Е.А. Избыточность в полупроводниковых интегральных микросхемах памяти. Мн.: Навука і тэхніка, 1995. 262 с.
9. Мак-Вильмс Ф.Дж., Слоэн Н.Дж. Теория кодов, исправляющих ошибки: Пер. с англ. М.: Связь, 1979. 744 с.