# ИНФОРМАТИКА

*Рашит Тагирович ФАЙЗУЛЛИН —*
*проректор по информатизации*
*Омского государственного технического*
*университета,*
*доктор технических наук, профессор*
*frt@omgtu.ru*
*Ильдар Рашитович ФАЙЗУЛЛИН —*
*аспирант*
*Омского государственного университета*
*blackildar@list.ru*

**УДК 519.7**

## *АЛГОРИТМ КОДИРОВАНИЯ И ПЕРЕДАЧИ ИНФОРМАЦИИ, БАЗИРУЮЩИЙСЯ НА СТЕГАНОГРАФИЧЕСКОМ ПОДХОДЕ*

## *DATA CODE AND COMPRESS ALGORITHM BASED ON STEGANOGRAPHY APPROACH*

*АННОТАЦИЯ. Предложен аппаратно-эффективный алгоритм кодирования и сжатия данных, основанный на применении стеганографической ключевой информации.*

*SUMMARY. The article presents hardware effective algorithm for data code and compression based on stenography key information.*

*КЛЮЧЕВЫЕ СЛОВА. Маркер адреса устройства, кодирование, сжатие.*
*KEY WORDS. Marker for hardware address, coding, compression.*

Постоянное увеличение объемов передаваемых данных и пропускной способности каналов связи приводит к качественному изменению подходов к проблеме передачи и кодирования информации. Стеганография [1], скрытые каналы [2], [3], сегментирование данных [3], накопительные водяные цифровые знаки [4], из экзотических инструментов становятся не только перспективными способами практической защиты информации, но и ее передачи. Принципиальным элементом вышеприведенных подходов является необходимость распознавания самого факта передачи данных доверенным пользователем, и естественным образом встает вопрос о возможности привлечения алгоритмов и методов, разработанных для защиты информации, непосредственно к задачам кодирования.

Рассмотрим задачу определения адреса пакета данных, передаваемого через шину данных, состоящую из $KN$ каналов. Допустим, что к шине подключены несколько устройств параллельной обработки информации и желательно, чтобы всего за несколько тактов работы, вне зависимости от величины $K$ и $N$, любое устройство распознавало факт адресации пакета именно к данному устройству. Такое очевидное решение, как выделение части бит пакета под биты адресации при увеличении числа устройств наталкивается на принципиальные трудности: адресное пространство устройств растет, т.к. адреса должны быть кодированы с гарантией от ошибки и соответственно растут размеры $K$ и $N$. Кроме того, такой подход совершенно не обеспечивает конфиденциальности передачи данных, что в настоящее время уже является хотя и не явным, но необходимым требованием при организации протоколов передачи данных.

Мы попытаемся предложить некую модификацию тривиального подхода, кодирующую адрес устройства в преобразованиях, производимых с содержимым пакета, результат которых, кроме адреса, одновременно содержит и дополнительную информацию. Первоначально будем считать пакет, состоящий из $KN$ бит маркером начала передачи данных. В работах [6], [7], [8] предлагаются различные алгоритмы встраивания маркера, но они малоэффективны, т.к. требуют обработки данных в режиме офф-лайн. Но оказывается, что можно существенно повысить эффективность данного подхода и в итоге осуществлять вставку и извлечение в режиме он-лайн.

Рассмотрим кортеж «байтов» $\{Q^1,..,Q^N\}$, где каждое $Q^j$ — это битовый вектор длины $K$, и соответствующую им последовательность бит $q^1,...,q^N$. Будем называть объединение этих кортежей мастер-ключом, известным только доверенным сторонам обмена данных.

Передающая сторона, или Алиса, генерирует кортеж байтов $\{X^1,...X^N\}$ с условиями:

$$X^j \geq Q^j, \text{ если } q^j = 0,$$
$$X^j < Q^j, \text{ если } q^j = 1, \qquad (1)$$

где неравенства понимаются как неравенства между целыми числами, содержащимися в соответствующих байтах, и маркируют начало сообщения операцией побитового сложения байтов $X^j, Q^j$:

$$Z^i = X^i \oplus Q^i \; i = 1, \ldots K. \qquad (2)$$

Отметим, что соотношения (1) и (2) можно рассматривать, как соответствия, а $q^1,...,q^N$ как условия, позволяющие выбрать единственный селектор из пересечения графиков этих соответствий.

Получатель сообщения, или Боб, складывает побитово байт $Z^j$ с байтом $Q^j$, получая в итоге $\overline{X}^j$. Если некоторая последовательность битов $\overline{X}^m,..,\overline{X}^{m+N-1}$ удовлетворяет условиям (1), то Боб делает вывод о том, что с вероятностью $1-2^{-N}$ эта последовательность байтов маркирует начало сообщения.

Но Алиса может специально генерировать некоторые из $X^j$ так, что условия (1) для некоторых из них не будут выполняться. Если число таких «ошибок» будет относительно мало, то Боб, проверяя входной поток, может сделать хорошо обоснованный вывод о наличии специально встроенных Алисой символов «ошибки». Боб должен выбрать из двух гипотез: он получает случайный поток данных или поток данных, созданный Алисой. Учитывая, что случайный поток данных, например, созданный с помощью генератора псевдослучайных чисел BBS, достаточно хорошо моделирует выход схемы Бернулли, то выделение кортежа $\overline{X}^m,..,\overline{X}^{m+N-1}$ не представляет большого труда, с учетом, например, того, что доверенное лицо проводит еще тест на открытый текст, для наиболее вероятного фрагмента потока данных.

В этом случае Боб может интерпретировать маркер как битовую строку $\overline{q}^j, j=1, ... N$, где нулями являются те из $\overline{q}^j$, которые совпадают с $q^j$, а единицами — места «ошибок». Длина слова в этом случае будет равна $N$, а «полезное пространство сообщений» $2^l$, где $l << N$.

Можно убрать это ограничение и «окаймлять» значимую информацию последовательностью «нулей», число которых заранее задано.

Насколько стойкой является предложенная схема, рассматриваемая как стеганографическая?

Предположим, что аналитик каким-то образом смог выделить $M$ кортежей: $Z^{js}, j=1, ..., N, s=1, ..., M$, отвечающих маркерам и ему известно, что в маркерах нет «ошибок». С учетом независимости блоков аналитик может рассматривать только один блок $j$ и отвечающие ему неизвестные: $q^j, Q^j, X^{j1}, ... X^{jM}$.

В этом случае мы получаем систему линейных уравнений в $Z_2$ и неравенств в $R$:

$$z^{js}{}_i = x^{js}{}_i \oplus Q^{js}{}_i \; i=1, ... K, s=1, ..., M,$$

$$(X^{js} - Q^j)(-1)^{q^j} \geq 0. \qquad (3)$$

Предполагая, что, например, $q^{js}=0$ получим:

$$(X^{js} - Q^{js}) \geq 0. \qquad (4)$$

Попытка полностью разрешить эту систему при $M=1$ обречена на неудачу, т.к. в сообщении не содержится никакой информации об используемой только один раз последовательности $q^1,...,q^N$ — это схема с одноразовым блокнотом.

Предположим, что $M=2$ и оценим число операций, необходимых для дешифрования в наиболее благоприятном для атакующего случае. Оказывается, что в этом случае верна лемма.

***Лемма.*** В наилучшем для атакующего случае, когда уже точно выделены два кортежа $Z^j$, трудоемкость определения кортежей (атака на обнаружение маркера) $Q^j$ и $q^j$ будет не меньше, чем $C(K)N2^K$.

Рассмотрим значения $Q^j$, которые возможно являются битами маркера без ошибок, очевидно, что их число оценивается величиной $2^K$.

Для каждого такого $Q^j$ однозначно определяется $X^j$ и проверяется, например, условие $(X^j - Q^j) \geq 0$. По результатам проверки $Q^j$ заносится в одно из множеств:

$$A = \{Q^j, X^j - Q^j \geq 0\} \text{ или } B = \{Q^j, X^j - Q^j < 0\}. \quad (5)$$

Выберем при $M = 2$ другой байт $Q^{j1}$ и так же образуем множества

$$A' = \{Q^j, X^{j1} - Q^j \geq 0\},$$

$$B' = \{Q^j, X^{j1} - Q^j < 0\}. \quad (6)$$

Предположим, что все пересечение $A \bigcap A'$ состоит только из одного $H^j$, а пересечение $B \bigcap B'$ пусто. В этом случае $H^j$ будет искомым байтом, а искомый бит $q^j = 0$.

Мы вынуждены решить системы два раза для каждого кортежа и константа $C$ очевидным образом как минимум пропорциональна $K$, длине байта. Другой подход, когда накапливаются $M$ кортежей, требует, кроме сравнимого числа операций, еще и экспоненциальной памяти для хранения данных.

Циклический сдвиг кортежа $q^1, ..., q^N$ и встраивание «ошибок» или сообщения решает эту проблему практически, но лемма по существу остается верной. Мы с вероятностью 0,25 попадаем в отрезок Голомба [9] длины больше чем единица.

Таким образом, гарантированную практическую стойкость на сегодняшний день, даже без учета операций записи и считывания из памяти, обеспечивает длина байта $K$, большая или равная 64.

Заметим, что реализация кодирования информации, или создание последовательности $X^1, ..., X^N$, может быть осуществлена инвертированием бит, например, применением операций И, ИЛИ с битами байта $Q$ и содержимым некоего регистра сдвига с линейной обратной связью. Если, например, $q^j = 0$, то некоторые нулевые биты $Q^j$ необходимо инвертировать на единичные (ИЛИ), если же $q^j = 1$, то наоборот, необходимо инвертировать единичные биты $Q^j$ (И).

Принимающая сторона может определить набор $q^1, ..., q^N$, пропуская его через регистр и тестируя байты, проверяя условие:

$$(X^{js} - Q^j)(-1)^{q^j} \geq 0, \quad (7)$$

например, с помощью релейной схемы, предложенной в [10].

Обратим внимание на то обстоятельство, что аппаратное шифрование и дешифрование происходит за $O(1)$ время, необходимое для прохождения сигнала через автомат, реализующий КНФ, а программная реализация требу-

ет $C(K)N$ числа операций, где функциональная зависимость, по крайней мере, линейная. Это позволяет сделать вывод о перспективности использования данного подхода в практике корпораций и неэффективности программных реализаций частными лицами.

Рассмотрим возможность использования в качестве сообщения самих $X^j$ без использования кортежа $Q^1$, ..., $Q^N$. Будем считать, что число $K$ четное, и подсчитаем для каждого $X^j$ количества $I_1$ и $I_2$ ненулевых бит в $X_1^j$ и в $X_2^j$, где $X^j = (X_1^j, X_2^j)$, т.е. $X^j$ — конкатенация, например, половин исходного битового вектора. Тогда вместо условий (1) мы можем записать:

$$I_1 \geq I_2, \text{ если } q^j = 0, \qquad (8)$$
$$I_1 < I_2, \text{ если } q^j = 1.$$

В этом случае перед отправкой сообщения каждый вектор $X^j$ суммируется по модулю 2 с битовым вектором, состоящим из единиц, или остается прежним, в зависимости от того удовлетворяет он (8) или нет. Проверку условия (8) можно организовать с помощью нейронной сети, которая состоит из двух суммирующих нейронов, объединенных в схему «победитель получает все».

Обратим внимание на то, что вывод о парадоксальности передачи данных с помощью ошибок в $q^1$, ..., $q^N$ «по верху» основной передачи данных $X^j$ неверен. Если передаваемый текст не обладает избыточностью, то информация о предварительном инвертировании с целью удовлетворения (8) теряется и в общем случае выигрыша мы не получаем. Кроме того, мы неявно используем тот факт, что $q^1$, ..., $q^N$ является маркером передачи выделенных блоков данных.

Но если у нас есть тест на открытый текст для каждого блока $j$, мы получаем дополнительную практическую процедуру сжатия информации. Например, если каждый блок уже архивирован, то применяя процедуру разархивирования, когда учитывается предварительное сложение по модулю два или нет, мы восстанавливаем текст однозначно.

## СПИСОК ЛИТЕРАТУРЫ

1. Грибунин В.Г., Оков И.Н., Туринцев И.В. Цифровая стеганография. М.: Солон-Пресс, 2002. 325 с.

2. Simmons, G. J. Subliminal Channels: Past and Present. European Trans. on Telecommunications, 4(4):459-473, Jul/Aug 1994.

3. Грушо А. А., Грушо Н. А., Тимонина Е. Е. Методы защиты информации от атак с помощью скрытых каналов и враждебных программно-аппаратных агентов в распределенных системах // Вестник РГГУ. 2009. № 10. С. 33-45.

4. Ренжин П.А., Файзуллин Р.Т. Методика защиты цифровых видеодоказательств от фальсификации встраиванием цифрового водяного знака // Тез. докл. «Научная сессия ТУСУР-2010». В 5 ч. Томск, 2010. Ч. 3. С. 191-192.

5. Blum, L., Blum, M., and Shub., M. A Simple Unpredictable Pseudo-Random Number Generator // SIAM Journal on Computing. Vol. 15. 1986. May. Pp. 364-383.

6. Файзуллин И.Р. Криптостеганографический алгоритм с использованием хэш-функции для маркировки начала сообщения // Тез. докл. «Научная сессия МИФИ-2007». Т. 16. Компьютерные науки. Информационные технологии. С. 149-150.

7. Delannay, D., Macq, B. A Method for Hiding Synchronization Marks in Scale and RotationResilient Watermarking Schemes, SPIE Electronic Imaging 2002, Security and Watermarking of Multimedia Contents IV, Int'l Soc. for Optical Eng. (SPIE). 2002. Vol. 4675. Pp. 548-554.

8. Lichtenauer, J., Setyawan, I., Lagendijk, R. Hiding correlation-based watermark templates using secret modulation. Security, Steganography, and Watermarking of Multimedia Contents. 2004. Vol. 5306. Pp. 501-512.

9. Golomb, S.W. Run-length encodings // IEEE Trans. Inf. Theor. 1996. IT-12. № 3. Pp. 399-401.

10. Дулькейт В.И. Сведение задач факторизации, дискретного логарифмирования и логарифмирования на эллиптической кривой к решению ассоциированных задач «Выполнимость» // Компьютерная оптика. 2010. Т. 34. № 1. С. 118-123.

***Александр Анатольевич ЗАХАРОВ —***
***зав. кафедрой информационной безопасности,***
***доктор технических наук, профессор***
***azaharov@utmn.ru***

***Ирина Гелиевна ЗАХАРОВА —***
***зав. кафедрой программного обеспечения,***
***доктор педагогических наук, профессор***
***izaharova@utmn.ru***
***Институт математики и компьютерных наук***
***Тюменский государственный университет***

**УДК 004.82**

# *ДИСКУРСИВНАЯ МЕТРИКА В ИНФОРМАЦИОННОМ ПРОСТРАНСТВЕ**

# *DISCOURSE METRICS IN INFORMATION SPACE*

*АННОТАЦИЯ. В статье предложен подход к определению качества web-ресурсов на основе дискурсивной метрики, которая учитывает как характеристики ресурса, так и динамический портрет пользователя.*

*SUMMARY. The article offers an approach to determining the quality of web-resources on the basis of discursive metrics which takes into account both the characteristics of the resource and dynamic portrait of the user.*

*КЛЮЧЕВЫЕ СЛОВА. Web-ресурсы, метрика, дискурс.*
*KEY WORDS. Web-resources, metrics, discourse.*

Современные исследования разнообразных проблем, связанных с информационными процессами, часто опираются на понятие «информационное пространство» — обозначающее особый феномен, возникший в условиях ин-

* Работа выполнена при финансовой поддержке Министерства образования и науки Российской Федерации в рамках Федеральной целевой программы «Научные и научно-педагогические кадры инновационной России» на 2009-2013 годы (ГК № 02.740.11.0594).