УДК 003.26+347.78

Н. П. Шутько
Белорусский государственный технологический университет

АЛГОРИТМЫ РЕАЛИЗАЦИИ МЕТОДОВ ТЕКСТОВОЙ СТЕГАНОГРАФИИ НА ОСНОВЕ МОДИФИКАЦИИ ПРОСТРАНСТВЕННО-ГЕОМЕТРИЧЕСКИХ И ЦВЕТОВЫХ ПАРАМЕТРОВ ТЕКСТА

Рассмотрены алгоритмические особенности реализации методов текстовой стеганографии для скрытой передачи данных и охраны прав интеллектуальной собственности. Методы основаны на модификации пространственно-геометрических (апрош, кернинг) и цветовых параметров символов текста. Скрытие данных производится не только в обычных, но и в специальных (мягкий перенос, разрыв строки и др.) символах и пробелах. Особенностью метода на основе модификации цвета является то, что процессы осаждения/извлечения информации осуществляются при сравнительном изменении/анализе цветовых параметров пар соседних символов, один из которых является базовым. Информация осаждается в соседний с базовым символ текста. Формальное описание преобразований осуществляется на основе математической модели двухключевой стеганографической системы: один ключ определяет метод осаждения, другие – иные преобразования осаждаемой информации и выбор элементов текста для их модификации. Методы и алгоритмы на основе модификации кернинга и апроша предусматривают изменение значения соответственно кернинга любых заданных дополнительным ключом кернинговых пар и апрошей в документе-контейнере. Методы характеризуются более высокой эффективностью в сравнении с другими методами текстовой стеганографии.

Ключевые слова: текстовая стеганография, алгоритм, математическая модель, параметры шрифта, авторское право.

N. P. Shut'ko
Belarusian State Technological University

THE ALGORITHMS OF REALIZATION OF TEXT STEGANOGRAPHY METHODS BASED ON THE MODIFICATION OF THE GEOMETRIC AND COLOR TEXT PARAMETERS

We consider the algorithmic features of implementation the text steganography methods to secure communication and protection of intellectual property rights. The methods are based on the modification of the space-geometrical (aprosh, kerning) and color settings of the text symbols. Hiding the data is done not only in the usual but also in special (hyphens, line break, etc.) characters and spaces. The peculiarity of the method on the basis of the color modification is that the precipitation/extraction information processes are carried out in a comparative change/analysis of the color parameters of pairs of adjacent characters, one of which is the basic. Information is deposited into the neighboring to the basic text symbol. A formal description of transformations is made on the basis of the mathematical model of two-key steganography system: one key determines the precipitation method, others – another transformation of the deposited information and a selection of text elements to modify them. Methods and algorithms based on the modification of kerning and aprosh provide changing the value of accordingly kerning of any given additional key kerning pairs and aprosh in the document container. The methods are characterized by higher efficiency in comparison with other methods of text steganography.

Key words: text steganography, mathematical model, algorithm, font parameters, copyright.

Введение. Развитие информационных технологий за последние десяток-полтора лет привело к тому, что значительная (и важнейшая) часть информации, относящейся к различным сторонам деятельности предприятий или организаций, теперь находится в электронном виде (в системах хранения данных). Наиболее характерным примером для академической среды являются репозитории электронных образовательных ресурсов. Поэтому особенную остроту приобретает проблема надежной защиты этих ресурсов, а также иных текстовых документов, программных кодов, баз данных от несанкционированного использования. Существуют различные способы для реализации такой защиты. К ним можно отнести, в частности, методы текстовой стеганографии, которые в последнее время становятся популярными. Информация, которая позволяет защитить права собственности на документ, скрывается (осаждается), наподобие цифровых водяных знаков, в документе (контейнере).

Скрывать такую информацию можно, используя различные элементы текста путем их незаметной модификации: например, междустрочного интервала, или интерлиньяжа (Line-Shift Coding), пробельного расстояния между словами (Word-Shift Coding). Особенности упомянутых методов проанализированы, например, в [1]. Их эффективность низка из-за сравнительно небольшой относительной части упомянутых элементов текста к общему числу его элементов.

Мы разработали ряд стеганографических методов, в основе которых лежит использование других пространственно-геометрических параметров элементов текста [2, 3]. Как известно, указанные параметры напрямую определяются шрифтом, на основе которого создан документ. Таких параметров достаточно много, что вызвано неоднородностью размеров шрифтовых знаков, сложившейся в результате его исторического развития. К числу основных относятся гарнитура, кегль, начертание и размеры очка (расстояние между верхней частью букв «Т» или «f» и нижней частью букв «р» или «g»). Кегль – это основа, все остальные пропорции шрифта напрямую зависят от него. Важнейший параметр шрифта – ширина символов. Здесь неоднородность максимальна (сравним, например, ширину символов «Щ» и «1»). Кроме ширины символов, строка текста формируется также, как было отмечено в характеристике метода Word-Shift Coding, пробельным материалом. Однако у пробельного материала есть родственные элементы и в пространстве литер, на так называемой кегельной площадке. Это апроши – пространство, намеренно оставленное с левой и правой сторон очка (отпечатываемого, видимого контура буквы), чтобы соседние буквы не налезали друг на друга, чтобы между ними всегда было расстояние.

И, наконец, последний пространственно-геометрический параметр – кернинг, свойственный так называемым кернинговым парам. Примерами кернинговых пар могут быть: ГА, ТА, АТА, ЬТ и т. п. (в текстах на основе кириллицы); AY, Av, T;, ff (на основе латиницы); ΘA, ΔO, λκ (на основе греческого алфавита).

Компьютерная графика «снабдила» видимые символы текста еще одним существенным параметром – цветом, который также может быть использован для осаждения тайной информации в текст по аналогии с известным методом графической стеганографии – LSB (Least Significant Bit – наименее значащий бит).

Далее рассмотрим алгоритмические особенности реализации методов, в которых осаждение информации производится путем модификации цвета символов, апроша и кернинга.

Элементы и описание стеганографической системы. Отметим несколько базовых понятий. Документ, который мы хотим защитить, называется *контейнером*, или файлом-контейнером *C*. Документ (текст), с помощью которого осуществляется защита путем его размещения (осаждения) в контейнере, или же который размещается для передачи, носит название *стегосообщения M*. Защищенный документ (контейнер с осажденным сообщением) называется *стеганоконтейнером S*.

В математической модели стеганографической системы пространство ключей (наподобие криптосистемы) образуют базовые методы осаждения/извлечения тайной информации (модификация цвета, апроша, кернинга или иные – основной ключ) и дополнительные преобразования, определяемые дополнительными ключами [4]. В соответствии с этим стеганографическую систему мы классифицируем как «двуключевую».

Формально модель задается следующим выражением:

$$\sum\nolimits_2 = (\mathbf{M}, \mathbf{C}, \mathbf{K}, \mathbf{K}^{\text{д}}, \mathbf{S}, \mathbf{F}, \mathbf{F}^{-1}). \qquad (1)$$

Параметрами в соответствии с приведенным соотношением являются: множество стегосообщений ($\mathbf{M}$), множество контейнеров ($\mathbf{C}$), множество ключей ($\mathbf{K}$) и дополнительных ключей ($\mathbf{K}^{\text{д}}$), множество стегоконтейнеров ($\mathbf{S}$), прямое ($\mathbf{F}$) и обратное преобразования ($\mathbf{F}^{-1}$). Прямое преобразование соответствует осаждению секретного сообщения в контейнер, обратное, в свою очередь, – извлечению.

Алгоритмы прямого стеганографического преобразования. Цвет пикселей, формирующих символы текста, пространственно-геометрические параметры текста можно изменить так, что это остается незаметным для других лиц в силу специфики человеческого зрения. Это, так сказать, психофизиологическая особенность методов.

Мы исходим из того, что между алгоритмами прямого преобразования (осаждение тайного сообщения в контейнер) и обратного преобразования (извлечение сообщения из контейнера) существует однозначное соответствие в виде функций $F \in \mathbf{F}$ и $F^{-1} \in \mathbf{F}^{-1}$. С учетом этого дальнейшее рассмотрение ограничим алгоритмами прямого преобразования.

Модификация цветовых параметров. В данном случае анализируемый метод определяется ключом K, $K \in \mathbf{K}$. Формально рассматриваем текстовый документ-контейнер как графический объект. В этом и других методах в качестве базового элемента контейнера, свойства которого модифицируются при осаждении

информации, выступает символ текста, включая пробел.

Используется цветовая модель **RGB**, в которой каждый цветовой канал задается 8-разрядным двоичным вектором либо соответствующим десятичным числом:

$$R, G, B \in \{0, 1\} \text{ либо } R, G, B \in \{0, 1, \ldots, 255\}.$$

Во втором случае часто говорят об интенсивности трех составляющих цвета.

Если текст-контейнер состоит из Z символов, то формально каждый из них может быть представлен как

$$z_t = \{r_t, g_t, b_t\} \in RGB,\ 1 \le t \le Z,$$

где t – порядковый номер символа в текстовом документе-контейнере.

Оригинальность метода состоит в том, что процессы осаждения/извлечения информации осуществляются при сравнительном изменении/анализе цветовых параметров пар соседних символов:

$$\{r_t, g_t, b_t\} \text{ и } \{r_{t+1}, g_{t+1}, b_{t+1}\}.$$

При этом цветовые координаты $\{r_t, g_t, b_t\}$ являются базой (центроидой) для осаждения/извлечения определенного знака («0» или «1») сообщения M. Изменение параметра соседнего символа оценивается по отношению к этому базовому.

Таким образом, если объем встраиваемого сообщения M_i равен n (пример: M_i = «Надя», $n = 4$) символов с пробелами исходного алфавита, анализ следует производить при преобразовании M_i в двоичную форму (поскольку цветовую модель мы рассматриваем на основе бинарных чисел): используем кодировку ASCII, в которой один символ представлен 8 битами ($Q = 8$), т. е. общая длина осаждаемого сообщения в бинарном виде N' составляет: $N' = Q \times n$ бит. Кроме того, необходимо иметь в виду количество необходимых символов N в контейнере C_j, которые потребуются для встраивания стегосообщения: легко понять, что Z должно быть не меньше $N = Q \times n \times 2$.

Выбор пар символов для осаждения информации определяется одним из дополнительных ключей: $K_1^{д} \in \mathbf{K}^{д}$. При этом $K_1^{д} = \{k_{11}^{д}, k_{10}^{д}\}$ и ключи $k_{11}^{д} = \{r_1, g_1, b_1\}$, $k_{10}^{д} = \{r_0, g_0, b_0\}$ используются соответственно при осаждении бинарных символов «1» и «0» (в z_{t+1}-й символ текста-контейнера):

$$z_{t+1} = z_t + k_{11}^{д} = \{r_t + r_1, g_t + g_1, b_t + b_1\}, \quad (2)$$

$$z_{t+1} = z_t + k_{10}^{д} = \{r_t + r_0, g_t + g_0, b_t + b_0\} \quad (3)$$

и $t = 1, \ldots, Z$, $Z \ge N$ – необходимое условие для успешного встраивания секретного сообщения в документ-контейнер, $z_t = \{r_t, g_t, b_t\} \in \{R, G, B\}$.

В процессе осаждения берутся в соответствии с $K_1^{д}$ два рядом стоящих символа текстового документа-контейнера: первый – базовый, второй – тот, в котором будет скрыт текст. Цвет символа, в котором будет производиться осаждение, формируется исходя из цвета символа-образца и заданного в настройках текстового процессора (через специальное программное приложение) смещения. По умолчанию это смещение добавляется к основному цвету. Если при этом значение выходит за допустимый диапазон – отнимается. Например, если цвет образца задан как {0, 200, 100}, а смещение задано как {100, 60, 50}, то результирующей будет цветовая координата {100, 140, 150}. Чтобы избежать ситуации, когда значение может выйти за оба диапазона, в полях для отклонения не рекомендуется задавать значения больше 128. В настройках указанного программного средства будет задаваться отклонение от основного цвета символа по трем цветовым каналам в соответствии с (2) и (3).

Например, возьмем в качестве контейнера C_j словосочетание «Текстовая стеганография» и разместим в нем секретное сообщение «Н» (M_i), используя описанный выше метод (рис. 1). Символы, выделенные полужирным начертанием, являются базовыми (центроидами), в подчеркнутые (входящие в пару) символы встраивается секретная информация: каждый разряд («0» или «1») встраивается в каждый подчеркнутый символ текста документа-контейнера. При этом для встраивания «1» используется ключ $k_{11}^{д}$, для «0», соответственно, – $k_{10}^{д}$.

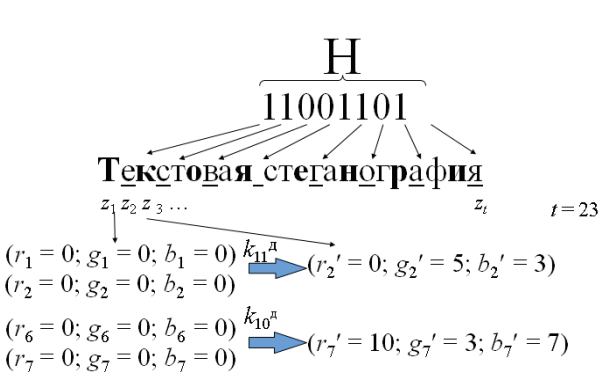

Рис. 1. Пояснение к методу модификации цветовых параметров

Алгоритм реализации метода модификации цвета символов представлен на рис. 2. Из алгоритма видно, что вначале идут параллельно три процесса. Это выбор текстового документа-контейнера, выбор секретного сообщения и выбор ключа.

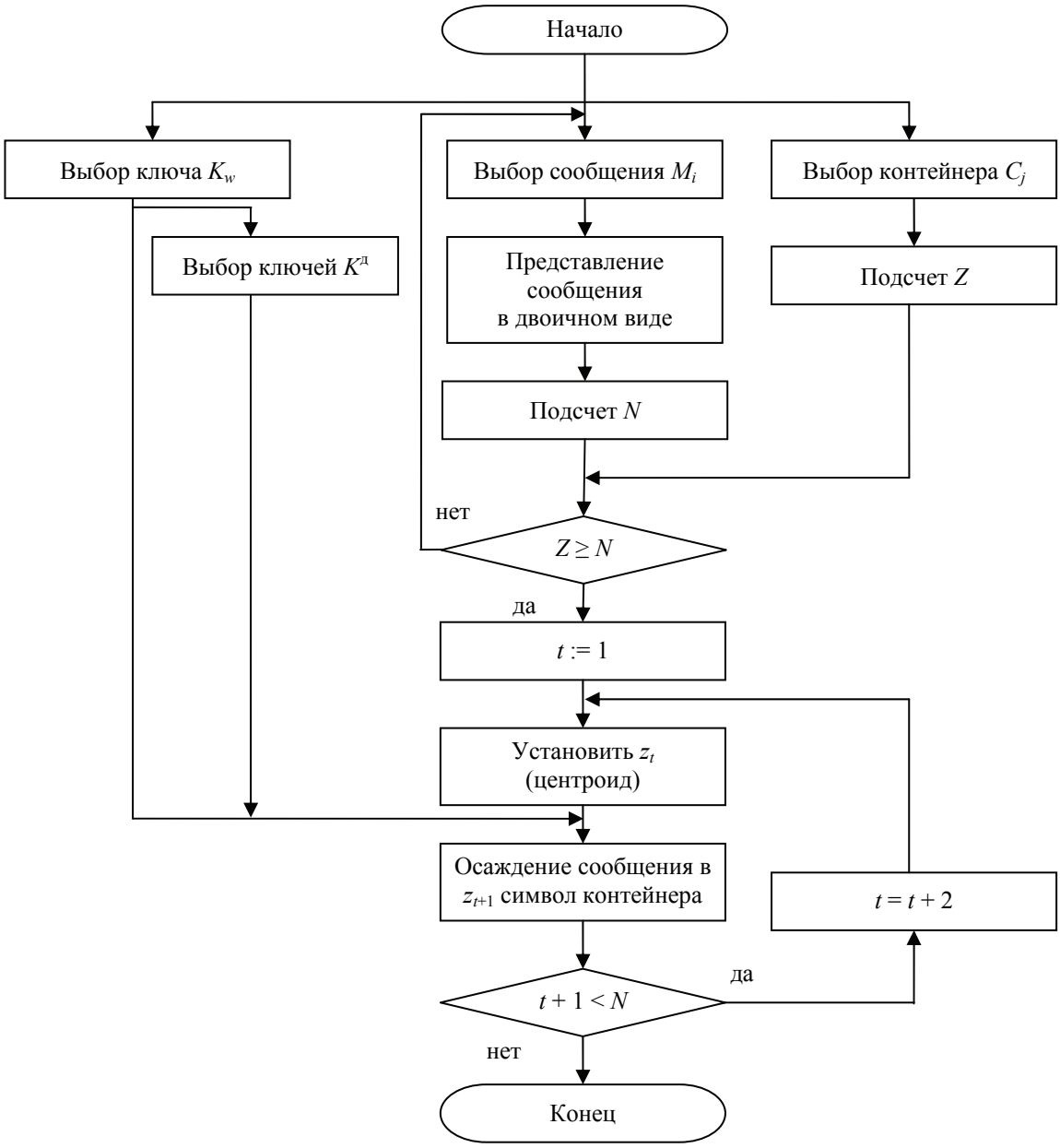

Рис. 2. Алгоритм метода изменения цвета символа

После того, как выбрано стегосообщение, оно переводится в двоичный вид. Далее производится вычисление N.

Например, необходимо встроить стегосообщение «Надя» (M_i) в текстовый документ-контейнер. Для этого:

– подсчитывается количество символов n в сообщении (если есть, то с пробелами); в нашем случае $n = 4$;

– M_i преобразуется в двоичный вид (используем кодировку ASCII):

11001101111000001110010011111111;

один символ в выбранной кодировке представлен 8 знаками ($Q = 8$);

– подсчитывается $N = Q \times n \times 2$; коэффициент 2 учитывает тот факт, что используется не один символ для осаждения информации, а два; в нашем случае $N = 64$, т. е. для встраивания секретного сообщения необходимо, чтобы текстовый документ-контейнер содержал как минимум 64 символа;

– в текстовом документе, который необходимо защитить (C_j), подсчитывается общее количество символов с пробелами Z и проверяется необходимое условие ($Z \geq N$). Далее устанавливается символ-образец.

Модификация апроша и кернинга. Сущность разработанных методов заключается в том, что незначительное изменение величины апроша и кернинга (в соответствии с ключом из множества K) относительно базового значения (доли пункта) не вызывает визуального восприятия уплотнения или разрежения групп символов.

Алгоритмы реализации рассматриваемых методов в большинстве своем схожи с методом изменения цветовых параметров. Однако имеются в каждом случае свои особенности.

Так, метод изменения кернинга можно реализовать двумя способами.

1. Осаждение информации производится за счет изменения значения кернинга любого символа в документе-контейнере.

2. Первоначально проводится анализ документа-контейнера на наличие в нем кернинговых пар. Существует таблица кернинговых пар для каждого семейства шрифтов. В ней приводятся те самые особые пары символов. Предполагается осаждать информацию за счет изменения значения кернинга именно между такими парами символов.

Поэтому разработанный алгоритм реализован по двум схемам. На рис. 3 приведена схема алгоритма для второго варианта реализации метода. Необходимо отметить, что в методах изменения апроша и кернинга производится подсчет только N' ($N' = Q \times n$) ввиду того, что осаждать информацию, используя дополнительный ключ $K^{д}$, можно в каждый символ документа-контейнера:

$$z_t = z_t + k_1^{д}. \quad (4)$$

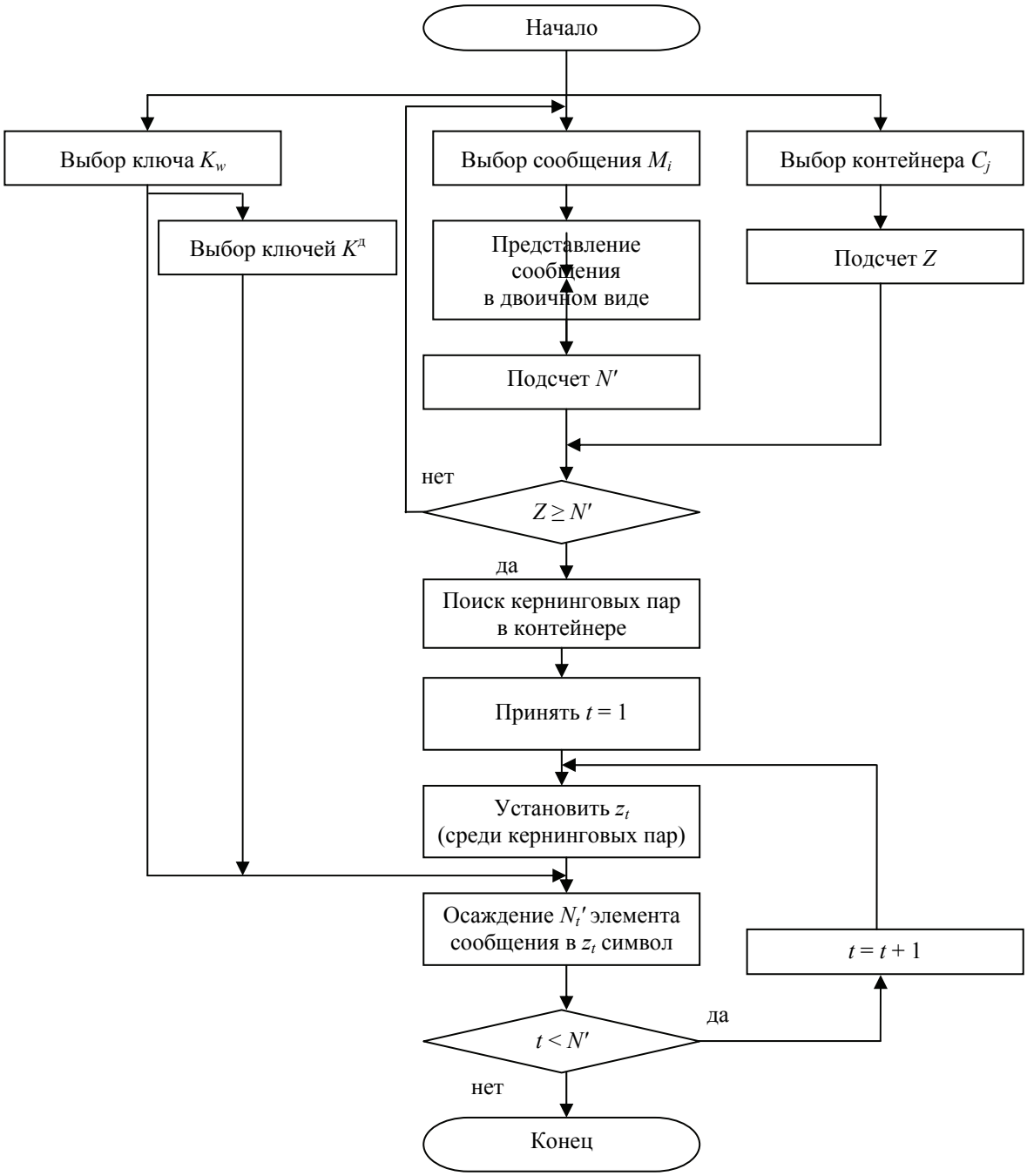

Рис. 3. Алгоритм метода изменения кернинга

Заключение. Разработанные и проанализированные на примерах алгоритмы реализации стеганографических методов на основе изменения пространственно-временных и цветовых параметров символов текста являются по уровню сложности сопоставимыми с алгоритмами методов модификации междустрочного интервала (Line-Shift Coding) и пробельного расстояния между словами (Word-Shift Coding), если принять, что в последних может применяться псевдослучайный выбор модифицируемых элементов.

Вместе с тем предлагаемые методы характеризуются большей эффективностью, поскольку число символов и их кернинговых пар в текстах всегда значительно превышает число строк или число слов. Необходимо также отметить, что стегознаками в методах изменения цвета символов, апроша и кернинга являются все символы документа, в том числе специальные знаки и символы.

Литература

1. Shutko N. Text steganography as an effective instrument of protection of the copyright on electronic document // New Electrical and Electronic Technologies and their Industrial Implementation: 8-th International Conference, Zakopane, Poland, june 18–21, 2013. Zakopane, 2013. P. 147.

2. Шутько Н. П. Защита авторских прав на электронные текстовые документы методами стеганографии // Труды БГТУ. 2013. № 6: Физ.-мат. науки и информатика. С. 131–134.

3. Shutko N. Text steganography method based on the change of font attributes // Printing future days: 6-th International Scientific Conference on Printing and Media Technology, Chemnitz, Germany, october 5–7, 2015. Chemnitz, 2015. P. 91.

4. Шутько Н. П., Романенко Д. М., Урбанович П. П. Математическая модель системы текстовой стеганографии на основе модификации пространственных и цветовых параметров символов текста // Труды БГТУ. 2015. № 6: Физ.-мат. науки и информатика. С. 152–156.

References

1. Shutko N. Text steganography as an effective instrument of protection of the copyright on electronic document. *New Electrical and Electronic Technologies and their Industrial Implementation: 8-th International Conference*. Zakopane, 2013, p. 147.

2. Shut'ko N. P. Copyright protection on the electronic text documents by methods of steganography. *Trudy BGTU* [Proceedings of BSTU], 2013, no. 6: Physical-mathematical sciences and informatics, pp. 131–134 (In Russian).

3. Shutko N. Text steganography method based on the change of font attributes. *Printing future days: 6-th International Scientific Conference on Printing and Media Technology*. Chemnitz, 2015, p. 91.

4. Shut'ko N. P., Romanenko D. M., Urbanovich P. P. Mathematical model of the text steganography on the base of modifying the spatial and color settings of text characters. *Trudy BGTU* [Proceedings of BSTU], 2015, no. 6: Physical-mathematical sciences and informatics, pp. 152–156 (In Russian).

Информация об авторе

Шутько Надежда Павловна – аспирант. Белорусский государственный технологический университет (220006, г. Минск, ул. Свердлова, 13а, Республика Беларусь). E-mail: NPCh@belstu.by

Information about the author

Shut'ko Nadezhda Pavlovna – PhD student. Belarusian State Technological University (13a, Sverdlova str., 220006, Minsk, Republic of Belarus). E-mail: NPCh@belstu.by

Поступила 07.03.2016