ОБРАБОТКА И ПЕРЕДАЧА ИНФОРМАЦИИ

УДК 003.26+347.78

Н. П. Шутько, Д. М. Романенко, П. П. Урбанович
Белорусский государственный технологический университет

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ СИСТЕМЫ ТЕКСТОВОЙ СТЕГАНОГРАФИИ НА ОСНОВЕ МОДИФИКАЦИИ ПРОСТРАНСТВЕННЫХ И ЦВЕТОВЫХ ПАРАМЕТРОВ СИМВОЛОВ ТЕКСТА

Приведено описание математической модели процессов, протекающих в стеганографической системе на основе модификации параметров символов текста. Основу математической модели составляют пространственные координаты (в качестве таковых используются апрош, кернинг, кегль шрифта) и цветовые параметры пикселей, формирующих растр изображения (текста). Модель основывается на теоретико-множественном определении функции. Тайная информация предназначена для защиты прав интеллектуальной собственности. Осаждение информации предусматривает изменение цветовых и пространственных координат символов текста. Рассмотрена сущность горизонтального и вертикального профилей текста, их построение по битовой карте изображения текста, а также изменение в результате встраивания тайной информации в апрош и кернинг. Рассматриваемая стеганографическая система представляет собой совокупность сообщений, контейнеров (или документов-контейнеров), двух ключей и преобразований, которые их связывают.

Ключевые слова: стеганография, авторское право, профиль текста, кегль, апрош, кернинг.

N. P. Shut'ko, D. M. Romanenko, P. P. Urbanovich
Belarusian State Technological University

MATHEMATICAL MODEL OF THE TEXT STEGANOGRAPHY ON THE BASE OF MODIFYING THE SPATIAL AND COLOR SETTINGS OF TEXT CHARACTERS

The description of the mathematical model of the processes occurring in the steganographic system based on the modification of the parameters of text characters is given. The basis of the mathematical model is the spatial coordinates (such as aprosh, kerning, font size) and color values of pixels forming the raster of the image (text). The model is based on a set-theoretic definition of the function. Secret information is intended to protect intellectual property rights. An embedding of the information provides the changing of the color and spatial coordinates of characters of the text. The essence of the horizontal and vertical profiles of the text, their construction by the bitmap of image of the text and changing as well as a result of embedding secret information in aprosh and kerning are viewed. The regarded steganographic system is a set of messages, containers (containers or documents), two keys and transformations by which they are linked.

Key words: steganography, copyright, text profile, font size, aprosh, kerning.

Введение. Все большие значение и роль приобретает цифровая форма данных. Активно осуществляется их хранение, передача и использование. В то же время возрастает и угроза цифрового пиратства. Каждый автор хочет быть уверен в защите своего труда. Именно поэтому в настоящее время все большую популярность приобретает стеганография. С помощью различных ее методов есть возможность оградить документ от несанкционированного копирования и распространения, тем самым защитить авторские права на интеллектуальную собственность.

В известных исследованиях авторов данной статьи [1–2], в других публикациях (например, [3]) обосновывается возможность использования некоторых пространственных или цветовых параметров символов текста, формируемого растром, для размещения тайной (авторской) информации, которая может использоваться в случае необходимости для доказательства права интеллектуальной собственности.

Идея, предложенная в [4] и состоящая в использовании массива пикселей изображения и соответствующих его профилей для количественных оценок параметров некоторых методов текстовой стеганографии, с нашей точки зрения, может стать хорошей методологической и методической основой и для новых методов. К последним относятся, например, методы, базирующиеся на изменении цветовых (RGB) и шрифтовых, или геометрических (размер, масштаб, гарнитура, начертание), параметров символов текстового документа.

Широкие графические, цветовые и пространственные возможности программного инструментария растровой графики позволяют достаточно просто изменять и анализировать геометрические параметры произвольного шрифта. Минимальной единицей растровой графики является пиксель (точка). Растровые изображения напоминают лист клетчатой бумаги, на котором любая клетка закрашена каким-либо цветом, образуя в совокупности рисунок (bitmap). Основными характеристиками растровой графики являются глубина цвета, разрешение и цветовая модель.

Чтобы лучше понять процессы, их взаимосвязь, необходимо математически их описать. Именно этому и посвящена статья.

Основная часть. Объектом исследования в данной работе являются стеганографические методы защиты прав интеллектуальной собственности на текстовые документы. Предметом – модели стеганографических процессов.

Кратко охарактеризуем основные из интересующих нас характеристик шрифта.

Кегль – это его вертикальный размер, измеряемый в пунктах (1 пункт равняется 0,376 мм) (рис. 1).

Рис. 1. Характеристики кегля шрифта

Апрош – расстояние между соседними буквами или другими шрифтовыми знаками. Изменение величины апроша относительно базового значения на небольшое расстояние (доли пункта) не вызывает визуального уплотнения или разрежения групп символов. Встраивание стегосообщения в контейнер на основе апроша производится путем изменения стандартного (базового) значения апроша до максимального (минимального) значения, зрительно не отличаемого от стандартного. Такое изменение производится с определенным шагом, каждому значению которого присваивается определенная комбинация бит.

Кернинг – избирательное изменение интервала между буквами в зависимости от их формы. Технология кернинга, появившаяся в полиграфии после внедрения фотонабора (а затем и компьютерного набора), включает подбор межбуквенных интервалов для конкретных пар букв с целью улучшения внешнего вида и удобочитаемости текста. Такой избирательный подбор позволяет компенсировать неравномерности визуальной плотности текста, получаемой при использовании стандартных апрошей для каждой буквы. Легко заметить, что расстояние между парами ИГ и ГА (рис. 2) выглядит разным, хотя формально оно одинаково. Для того чтобы между любыми двумя буквами расстояние было визуально одинаковым, и применяется кернинг (рис. 3).

Рис. 2. Текст без кернинга

Рис. 3. Текст с кернингом

Предлагаемая модель строится на основе следующих обозначений и положений:

пусть $\boldsymbol{M}$ – это конечное множество сообщений, которые могут быть тайно размещены в контейнере; $\boldsymbol{M} = \{M_1, M_2, \ldots, M_n\}$;

$\boldsymbol{B}$ – множество всех допустимых текстовых файлов-контейнеров; $\boldsymbol{B} = \{B_1, B_2, \ldots, B_p\}$, причем $p > n$;

$\boldsymbol{K}$ – множество всех допустимых ключей, под которыми будем понимать метод или алгоритм осаждения стегосообщения в контейнере; $\boldsymbol{K} = \{K_1, K_2, \ldots, K_z\}$.

Произвольное тайное сообщение M можно скрыть в контейнере B при использовании ключа K: $M \in \boldsymbol{M}$, $B \in \boldsymbol{B}$, $K \in \boldsymbol{K}$.

При этом получаем стегосообщение $S \in S = \{(M_1, B_1, K_1), (M_2, B_2, K_2), \ldots, (M_g, B_g, K_g)\} = S_1, S_2, \ldots, S_g$.

Дальнейшие рассуждения будем строить на базе основных понятий, которые будут сформулированы в виде определений.

Определение 1. Функцию $\boldsymbol{F}$, определенную на $\boldsymbol{M}\times\boldsymbol{B}\times\boldsymbol{K}$ со значениями в $\boldsymbol{S}$, будем отождествлять с осаждением или встраиванием сообщения $\boldsymbol{M}$ в контейнер $\boldsymbol{B}$ на основе использования пространственных или цветовых параметров элементов контейнера $\boldsymbol{B}$:

$$F: M\times B\times K \to S. \qquad (1)$$

Сообщением может являться любой текст, который необходимо тайно передать (встроить).

Контейнером называется файл (документ), в который происходит встраивание секретного сообщения.

Определение 2. Функцию $\boldsymbol{F^{-1}}$, определенную на $\boldsymbol{S\times K}$ со значениями в $\boldsymbol{M}$, будем отождествлять с извлечением тайного сообщения $\boldsymbol{M}$ из стегосообщения $\boldsymbol{S}$:

$$F^{-1}: S\times K \to M. \qquad (2)$$

Определение 3. Коллизией стеганографического преобразования (или пересечением) будем называть ситуацию, при которой

$$(M_a, B_a, K_a) = (M_b, B_b, K_b), \qquad (3)$$

причем $M_a \neq M_b$, $B_a \neq B_b$, $K_a \neq K_b$.

Имеется в виду, что может возникнуть такая ситуация, при которой после извлечения с помощью различных ключей из двух различных контейнеров секретных сообщений последние могут оказаться идентичными.

Определение 4. Стеганографической системой будем называть совокупность сообщений, контейнеров (или документов-контейнеров), ключей и преобразований, которые их связывают:

$$\sum = (M, B, K, \acute{S}, F, F^{-1}). \qquad (4)$$

Определение 5. Дополнительным ключом $\boldsymbol{K_д}$ стеганографической системы будем считать конкретное секретное значение набора параметров криптографического алгоритма, используемое для зашифрования ($E_{K_д}(M)$) и расшифрования ($D_{K_д}(S)$) сообщения (или, например, для помехоустойчивого кодирования/декодирования) соответственно при осаждении и извлечении; $K_д \in \boldsymbol{K_д} = \{K_{д1}, K_{д2}, \ldots, K_{дr}\}$.

Таким образом, стеганографические преобразования в общем виде описываются соотношениями

$$F: M\times B\times K_д\times K \to \acute{S}; \qquad (5)$$

$$F^{-1}: \acute{S}\times K\times K_д \to M; \qquad (6)$$

$$\sum = (M, B, K, K_д, S, F, F^{-1}). \qquad (7)$$

Такой вид системы будем называть двухключевой (заметим, в последнем случае вместо S используется $\acute{S}$).

Определение 6. Электронный документ-контейнер $\boldsymbol{B}$ будем представлять через дискретную функцию $\boldsymbol{f(x, y)}$, которая определяет координату для каждого пикселя изображения в двумерном пространстве (или массиве) $\boldsymbol{A}$; $\boldsymbol{x = 0, 1, \ldots, w}$; $\boldsymbol{y = 0, 1, \ldots, l}$.

Значение функции $f(x, y) \in \{0, 1\}$ – для монохромного или черно-белого изображения и $f(x, y) \in \{R, G, B\}$, где R, G, B – 8-битовые бинарные коды, определяющие спектр (цвет) каждого из каналов формирования изображения в так называемой аддитивной цветовой модели.

Как известно, считывание информации в растровой графике происходит по битовой карте. Рассмотрим на примере битовую карту двух строчек текста (рис. 4).

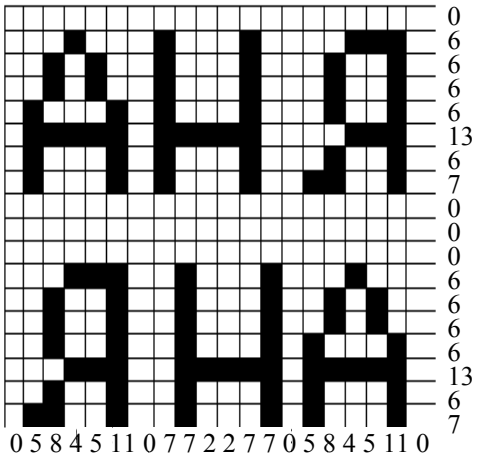

Рис. 4. Пример битовой карты

Определение 7. *Профилем* будем называть проекцию массива $\boldsymbol{A}$ или фрагмента этого массива, значения элементов которых определены дискретной функцией $\boldsymbol{f(x, y)} \in \boldsymbol{\{0, 1\}}$ для $0 \le x \le w$ и $0 \le y \le l$, на одну из осей – x или y.

Для анализа текста необходимо построить и проанализировать его горизонтальный и вертикальный профили. Сравнение профиля исходного текста с профилем переданного текста позволяет извлечь осажденную информацию.

Рассмотрим на примере построение горизонтального (рис. 5) и вертикального (рис. 6) профилей текста, указанного выше на битовой карте (рис. 4). Значения по осям – сумма закрашенных пикселей в битовой карте.

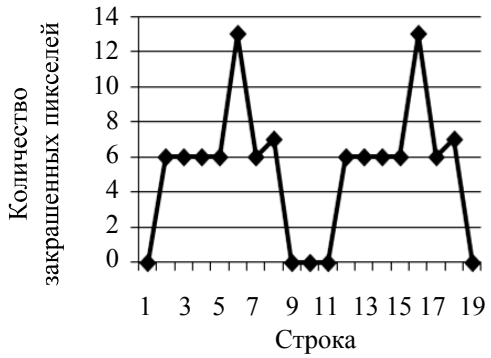

Рис. 5. Горизонтальный профиль

Горизонтальный профиль состоит из различных «вершин» и «впадин». «Вершины» соответствуют горизонтальным линиям сканирования вдоль линии текста, а «впадины» – межстроковому пространству. Ширина каждой

вершины в горизонтальном профиле соответствует «высоте тела» символов на текстовой линии; это около 41 пикселя при размере шрифта в 10 пунктов (т. е. 10/72 дюймов высоты символа). Рис. 5 и 6 показывают горизонтальный и вертикальный профили двух линий текста, содержащих по одному слову в каждой.

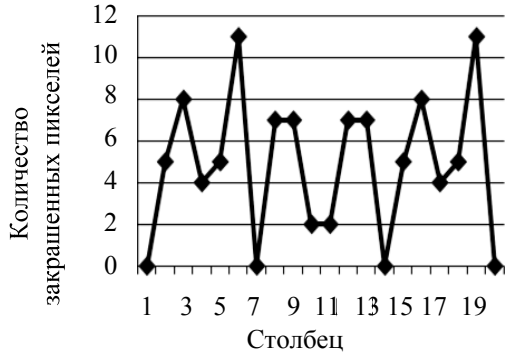

Рис. 6. Вертикальный профиль

Профиль содержит информацию об относительных положениях текста в изображении. При визуальном анализе профилей текста можно выявить измененные характеристики шрифта (например, апрош). Ниже приведены вертикальный и горизонтальный профили одной строки текста с одним словом в ней (рис. 7, 8).

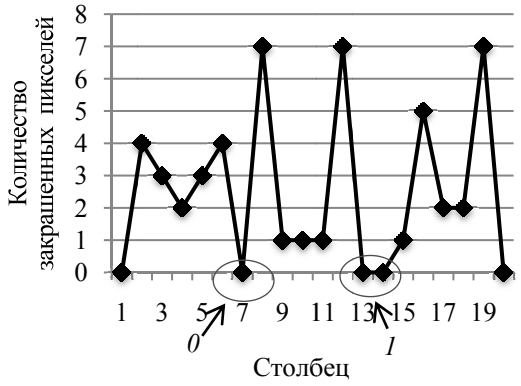

Рис. 7. Вертикальный профиль одной строки

Как видно из рис. 7, интервал между буквами различный, т. е. изменен апрош. Будем считать, что неизмененный апрош будет соответствовать «0», а измененный – «1». Таким образом, в данной строке текста содержится сообщение «01».

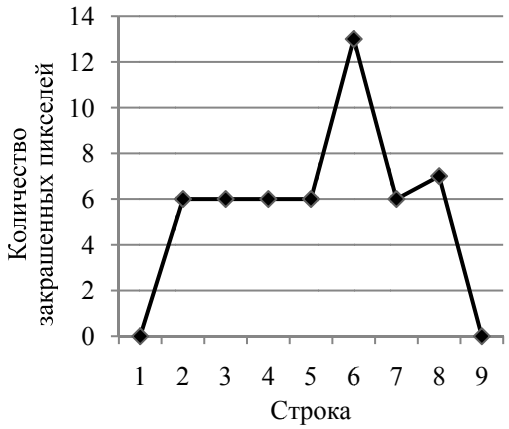

Рис. 8. Горизонтальный профиль одной строки

Заключение. Рассмотренная математическая модель основывается на теоретико-множественном определении функции.

Поскольку функция ***f*(*x*, *y*)** является дискретной, и такой же характер носит зависимость $M(x, y)$ при использовании ключа для зашифрования сообщения, то стегосообщение можно представить как результат аддитивного процесса: $\acute{S}(x, y) = f(x, y) + M(x, y)$.

В соответствии с этим извлечение информации состоит в анализе дискретной функции $\acute{S}(x, y)$, а также образующих ее горизонтального и вертикального профилей: $h(y) = \sum f(x, y)$ и $v(x) = \sum f(x, y)$.

Рассматриваемый тип системы стегопреобразования классифицируем как «двухключевая стегосистема»: один ключ определяет алгоритм осаждения/извлечения тайной (авторской) информации, другой – выбор символов текста для их последующей модификации.

Литература

1. Шутько Н. П. Особенности и формальное описание процесса осаждения секретной информации в текстовые документы на основе стеганографии // Труды БГТУ. 2014. № 6: Физ.-мат. науки и информатика. С. 121–124.

2. 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 Int. Konf., Zakopane, Poland, June 18–21, 2013. Zakopane, 2013. P. 147.

3. Конахович Г. Ф., Пузыренко А. Ю. Компьютерная стеганография. Киев: МК-Пресс, 2006. 288 с.

4. Document Marking and Identification using Both Line and Word Shifting / S. H. Low [et al.]. Boston: Infocom, 1995. 8 p.

References

1. Shut'ko N. P. Peculiarities and formal description of the embedding process of the secret information in text documents based on steganography. *Trudy BGTU* [Proceedings of BSTU], 2014, no. 6: Physical-mathematical sciences and informatics, pp. 121−124 (In Russian).

2. 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 Int. Konf*. Zakopane, 2013, p. 147.

3. Konakhovich G. F., Puzyrenko A. U. *Komp'yuternaya steganografiya* [Computer steganography]. Kiev: MK-Press Publ., 2006. 288 p.

4. Low S. H., Maxemchuk N. F., Brassil J. T., O'Gorman L. Document Marking and Identification using Both Line and Word Shifting. Boston, Infocom, 1995. 8 p.

Информация об авторах

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

Романенко Дмитрий Михайлович – кандидат технических наук, доцент, заведующий кафедрой информатики и компьютерной графики. Белорусский государственный технологический университет (220006, г. Минск, ул. Свердлова, 13а, Республика Беларусь). E-mail: rdm@belstu.by

Урбанович Павел Павлович – доктор технических наук, профессор, заведующий кафедрой информационных систем и технологий. Белорусский государственный технологический университет (220006, г. Минск, ул. Свердлова, 13а, Республика Беларусь). E-mail: p.urbanovich@belstu.by

Information about the authors

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

Romanenko Dmitriy Mikhaylovich – Ph. D. (Engineering), Assistant Professor, Head of the Department of Informatics and Computer Graphics. Belarusian State Technological University (13a, Sverdlova str., 220006, Minsk, Republic of Belarus). E-mail: rdm@belstu.by

Urbanovich Pavel Pavlovich – D. Sc. (Engineering), Professor, Head of the Department of Information Systems and Technologies. Belarusian State Technological University (13a, Sverdlova str., 220006, Minsk, Republic of Belarus). E-mail: p.urbanovich@belstu.by

Поступила 12.03.2015